

Fernando Celestino Moreira Vacari

Cristalização e resolução de estrutura das proteínas *Canavalia*
gladiata lectin (CGL) e *Canavalia marítima lectin* (CML)
complexadas ao açúcar manose 1-6 manose

São José do Rio Preto
2010

Fernando Celestino Moreira Vacari

Cristalização e resolução de estrutura das proteínas *Canavalia*
gladiata lectin (CGL) e *Canavalia marítima lectin* (CML)
complexadas ao açúcar manose 1-6 manose

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular, Área de Concentração - Física, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

Orientador: Prof. Dr. Valmir Fadel

São José do Rio Preto
2010

Vacari, Fernando Celestino Moreira
Cristalização e resolução de estrutura das proteínas
Canavalia gladiata lectin (CGL) e *Canavalia marítima lectin* (CML)
complexada ao açúcar manose 1-6 manose / Fernando Celestino
Moreira Vacari. - São José do Rio Preto: [s.n.], 2010.
79 f. : il. ; 30 cm.

Orientador: Valmir Fadel
Dissertação (mestrado) - Universidade Estadual Paulista, Instituto
de Biociências, Letras e Ciências Exatas

1. *Canavalia gladiata lectin*. 2. *Canavalia marítima lectin*. 3.
Cristalização. 4. Proteínas – Estrutura. I. Fadel, Valmir. II.
Universidade Estadual Paulista, Instituto de Biociências, Letras e
Ciências Exatas. III. Título.

CDU – 577.112

Fernando Celestino Moreira Vacari

Cristalização e resolução de estrutura da proteínas Canavalia
gladiata lectin (CGL) e Canavalia marítima lectin (CML)
complexadas ao açúcar manose 1-6 manose

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular, Área de Concentração - Física, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

Banca Examinadora

Prof. Dr. Valmir Fadel
UNESP – São José do Rio Preto
Orientador

Prof^a. Dr^a. Fernanda Canduri
Instituto de Química / Universidade de São Paulo

Prof. Dr. Mario Tyago Murakami
Centro de Biologia Molecular Estrutural / Associação Brasileira de Tecnologia de Luz Sincrotron

São José do Rio Preto
29/janeiro/2010

DEDICATÓRIA

Ao eterno amor da minha vida,
dedicada, companheira, ponto de
apoio nos momentos mais difíceis
e principal cúmplice na realização
deste trabalho.

AGRADECIMENTOS

Sou muito grato, primeiramente, a Deus, por ter me concedido a vida e ter sido, sempre, a luz do meu caminho.

Aos meus pais, por serem a base da minha estrutura familiar, refúgio cercado de amor e carinho em que sempre me abriguei e necessários na condução da minha formação enquanto pessoa, filho, amigo, estudante, profissional e cidadão.

Quero agradecer, imensamente, a minha instituição de ensino, Ibilce/UNESP, e todos os que nela colaboram, por ter me proporcionado tantas vivências inigualáveis, alegrias e pelas oportunidades de crescimento e amadurecimento durante esses quatro anos.

Ao Prof Dr Valmir Fadel, pela importante condução e orientação neste trabalho.

Ao meu amigo Diego por me ensinar os primeiros passos da cristalografia, com paciência e sabedoria.

Enfim, a todos que colaboraram, direta ou indiretamente, na força, incentivo e torcida para a finalização deste momento tão especial, com a conclusão deste trabalho e do tão sonhado Título de Mestre numa instituição pública e de qualidade de ensino.

RESUMO

Este trabalho teve por objetivo cristalizar e resolver tridimensionalmente as estruturas das proteínas *Canavalia gladiata lectin* (CGL) e *Canavalia marítima lectin* (CML), ambas complexadas com o açúcar manose 1-6 manose, encontradas em sementes de leguminosas. Para o processo de cristalização foi utilizado um kit de cristalização, contendo 96 soluções previstas pelo método de cristalização da matriz esparsa, denominado **Screen Index** (Hampton Research). Para o processo de resolução de estrutura foi utilizado diversos métodos computacionais dentre eles, o programa CCP4 (Collaborative Computational Project n° 4). Já com a estrutura resolvida, pode-se observar o sítio de ligação da proteína e identificar quais aminoácidos fazem parte do mesmo; calcular o RMSD médio entre as estruturas nativa e complexada com o açúcar manose 1-6 manose; comparar os sítios de ligação das proteínas CGL e CML complexadas com os açúcares man 1-2 man, man 1-3 man, man 1-4 man e man 1-6 man. Enfim, o trabalho colaborou para que duas novas estruturas fossem depositadas no banco de dados PDB (Protein Data Bank), para que futuros pesquisadores possam realizar estudos utilizando essas estruturas já resolvidas.

Palavras-chave: *Canavalia gladiata lectin* e *Canavalia marítima lectin*.

ABSTRACT

*This study aimed to crystallize and solve the three-dimensional structures of proteins Canavalia gladiata lectin (CGL) e Canavalia maritima lectin (CML), both complexed with sugar manose 1-6 manose, found in legume seed. For the crystallization process was used a kit of crystallization, containing 96 solutions provided by the method of crystallization of the sparse matrix, called **Screen Index** (Hampton Research). For the process of resolution structure was used several computation methods among them, the program CCP4 (Collaborative Computational Project n° 4). Now with the structure resolved, can observe the binding siti of protein and amino acids to identify which part of the same; calculate the avarage RMSD between the native and complexed structures with the sugar manose 1-6 manose; compare the binding sites of CGL and CML proteins complexed with sugars man 1-2 man, man 1-3 man, man 1-4 man and man 1-6 man. Finally, the work helped the two new structures were deposited in the PDB database, for future researchers to conduct studies using these structures already solved.*

Keywords: Canavalia gladiata lectin and Canavalia marítima lectin.

LISTA DE ILUSTRAÇÕES

Figura 1	Representação esquemática de merolectina (Peumans & Van Damme <i>et al.</i> , 1998).....	2
Figura 2	Representação esquemática de Hololectina (Peumans & Van Damme <i>et al.</i> , 1998).....	3
Figura 3	Representação esquemática de quimerolectina (Peumans & Van Damme <i>et al.</i> , 1998).....	4
Figura 4	Representação esquemática de Superlectina (Peumans & Van Damme <i>et al.</i> , 1998).....	5
Figura 5	Lectina presente na semente de <i>Parkia platycephala</i> . Essa lectina não se enquadra especificamente em nenhuma das classificações propostas por Peumans & Van Damme (Cavada <i>et al.</i> , 2006).....	6
Figura 6	Estrutura da lectina de leguminosa CGL man 1-4 man (PDB:2p2k).....	9
Figura 7	Redes de pontes de hidrogênio (presentes na lectina LoLI) formadas entre seis resíduos de aminoácidos e um açúcar simples. O resíduo aromático (phe ¹²³) cria uma interação hidrofóbica com o anel piranose do açúcar.....	11
Figura 8	Representação esquemática de um diagrama de solubilidade para uma macromolécula quando a concentração do agente cristalizante é alterada.....	16
Figura 9	Representação esquemática de (a) gota suspensa (hanging drop), (b) gota sentada (sitting drop) e (c) gota sanduíche (sandwich drop).....	18
Figura 10	Método da gota suspensa (hanging drop) para a cristalização de proteínas. O equilíbrio é alcançado pela difusão de vapor entre a gota e a solução precipitante.....	19

Figura 11	As celas marcadas com a letra P são primitivas, e as com letra M são múltiplas ou centradas.....	21
Figura 12	Representação de uma cela unitária geral. (a) Cela unitária com eixos a , b e c e ângulos α , β e γ situada na origem O . (b) A rede cristalina é obtida com o arranjo tridimensional das celas unitárias. (c) Qualquer cela unitária do cristal está relacionada com a cela da origem por uma operação de translação T	22
Figura 13	Os 14 retículos de Bravais. Eles são formados a partir de sete sistemas cristalinos: cúbico, tetragonal, ortorrômbico, monoclínico, triclínico, trigonal e hexagonal.....	24
Figura 14	Planos cristalinos com índices (<i>hkl</i>).....	27
Figura 15	Uma rede cristalina bidimensional ilustrando os índices de Miller para certo conjunto de “planos”. (a) Rede cristalina com uma cela unitária situada na origem O e três “planos” cruzando os eixos a e b . (b) As três famílias de “planos” são obtidas com a repetição dos “planos” originais.....	28
Figura 16	Espectro eletromagnético.....	29
Figura 17	Esquema da lei de Bragg.....	33
Figura 18	Gotas contendo cristais da proteína CML em a) e cristais da proteína CGL em b).....	41
Figura 19	Padrões de difração da proteína CML complexada com o açúcar man 1-6 man em a) e padrões de difração da proteína CGL complexada com o açúcar man 1-6 man em b).....	42
Figura 20	Alinhamento das estruturas primárias da CML e da CGL com 99% de identidade.....	46
Figura 21.a	Diagrama do dímero da lectina CML na perspectiva front face e back face respectivamente. Programa utilizado Pymol (DELANO, 2004). Representação Ribbon Cartoon.....	51

Figura 21.b	Diagrama do dímero da lectina CGL na perspectiva front face e back face respectivamente. Programa utilizado Pymol (DELANO, 2004). Representação Ribbon Cartoon.....	51
Figura 22	Quatros loops localizados na parte superior dos monômeros da lectinas CML e CML respectivamente...	52
Figura 23	Sítio de ligação do açúcar no monômero A da lectina CML.....	53
Figura 24	Sítio de ligação do açúcar no monômero B da lectina CML.....	55
Figura 25	Sítio de ligação do açúcar no monômero A da lectina CGL.....	57
Figura 26	Sítio de ligação do açúcar no monômero B da lectina CGL.....	59
Figura 27	Distância entre o átomo de Cálcio e Manganês presente no interior da proteína CML. Gerado pelo programa XtalView (MCREE, 1999).....	61
Figura 28	Diagrama do sítio de ligação dos átomos de Cálcio e Manganês da lectina CML. Programa utilizado Pymol (DELANO, 2004).....	62
Figura 29	Medida da distância do resíduo Asp ²⁰⁸ com a manose 1-6 manose da lectina CML gerada pelo programa XtalView (MCREE, 1999). Figura gerada pelo programa Pymol (DELANO, 2004).a) Monômero A. b) Monômero B.....	63
Figura 30	Medida da distância do resíduo Asp ²⁰⁸ com a manose 1-6 manose da lectina CGL gerada pelo programa XtalView (MCREE, 1999). Figura gerada pelo programa Pymol (DELANO, 2004). a) Monômero A. b) Monômero B.....	63
Figura 31.a	Diagrama de Ramachandran Φ - Ψ para a estrutura CML man 1-6 man. Imagem gerada utilizando o webserve RAMPAGE	65

Figura 31.b	Diagrama de Ramachandran Φ - Ψ para a estrutura CGL man 1-6 man. . Imagem gerada utilizando o webserve RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php)	65
Figura 32.a	Cálculo do RMSD do dímero da lectina CML (calculado para todos os átomos da estrutura) utilizando o programa Profit. Gráfico gerado pelo programa Origin...	67
Figura 32.b	Cálculo do RMSD do dímero da lectina CGL (calculado para todos os átomos da estrutura) utilizando o programa Profit. Gráfico gerado pelo programa Origin...	67
Figura 33.a	Sobreposição dos resíduos com o valor de RMSD relativamente altos da proteína CML nativa (em verde claro) e CML complexada com o açúcar man 1-6 man (em azul claro).....	68
Figura 33.b	Sobreposição dos resíduos com o valor de RMSD relativamente altos da proteína CGL nativa (em verde claro) e CGL complexada com o açúcar man 1-6 man (em azul claro).....	68
Figura 34.a	Cálculo do RMSD do dímero da lectina CML (calculado para os átomos da cadeia principal) utilizando o programa Profit. Gráfico gerado pelo programa Origin...	69
Figura 34.b	Cálculo do RMSD do dímero da lectina CGL (calculado para os átomos da cadeia principal) utilizando o programa Profit. Gráfico gerado pelo programa Origin...	70
Figura 35.a	Sobreposição dos resíduos do sítio ativo das proteínas CML nativa (em verde claro) e CML complexada com o açúcar man 1-6 man (em azul claro). a) Monômero A. b) Monômero B. Programa utilizado Pymol (DELANO, 2004).....	71
Figura 35.b	Sobreposição dos resíduos do sítio ativo das proteínas CGL nativa (em verde claro) e CGL complexada com o açúcar man 1-6 man (em azul claro). a) Monômero A. b) Monômero B. Programa utilizado Pymol	71

Figura 36 Comparação entre os sítios de ligação da lectina CGL 73
complexada aos açucares: a) man 1-2 man, b) man 1-
3 man, c) man 1-4 man e d) man 1-6 man.....

LISTA DE TABELAS

Tabela 1	Sistemas Cristalinos.....	23
Tabela 2	Estatística do processamento de dados de difração de raios X da lectina CML e da lectina CGL.....	43
Tabela 3.a	Estatística detalhada por faixa de resolução da lectina CML.....	44
Tabela 3.b	Estatística detalhada por faixa de resolução da lectina CGL.....	45
Tabela 4.a	Ângulos de rotação, distâncias de translação, coeficiente de correlação (CC) e R_{factor} obtidos na substituição molecular do cristal da lectina CML.....	47
Tabela 4.b	Ângulos de rotação, distâncias de translação, coeficiente de correlação (CC) e R_{factor} obtidos na substituição molecular do cristal da lectina CGL.....	47
Tabela 5.a	Ângulos de rotação, distâncias de translação, coeficiente de correlação (CC) e R_{factor} obtidos para quatro grupos espaciais do sistema ortorrômbico (CML).....	48
Tabela 5.b	Ângulos de rotação, distâncias de translação, coeficiente de correlação (CC) e R_{factor} obtidos para quatro grupos espaciais do sistema ortorrômbico (CGL).....	48
Tabela 6	Estatística do refinamento cristalográfico da proteína CML e da Proteína CGL.....	50
Tabela 7	Distâncias das possíveis pontes de hidrogênio presentes no sítio de ligação do Monômero A da lectina CML man 1-6 man. Calculado pelo programa XtalView (MCREE, 1999).....	54
Tabela 8	Distâncias das possíveis pontes de hidrogênio presentes no sítio de ligação do Monômero B da lectina CML man 1-6 man. Calculado pelo programa	56

	XtalView (MCREE, 1999).....	
Tabela 9	Distâncias das possíveis pontes de hidrogênio presentes no sítio de ligação do Monômero A da lectina CGL man 1-6 man. Calculado pelo programa XtalView (MCREE, 1999).....	58
Tabela 10	Distâncias das possíveis pontes de hidrogênio presentes no sítio de ligação do Monômero B da lectina CGL man 1-6 man. Calculado pelo programa XtalView (MCREE, 1999).....	60

SUMÁRIO

1	INTRODUÇÃO	1
1.1	LECTINAS	1
1.2	FAMILIAS EVOLUTIVAS	6
1.3	LECTINAS DE LEGUMINOSAS	8
2	OBJETIVOS	13
3	MATERIAIS E MÉTODOS	14
3.1	CRISTALIZAÇÃO DE MACROMOLÉCULAS BIOLÓGICAS	14
3.2	MÉTODOS DE CRISTALIZAÇÃO DE MACROMOLÉCULAS BIOLÓGICAS	17
3.3	CRISTAIS	20
3.4	ESTRUTURA CRISTALINA	20
3.5	ESPAÇO RECÍPROCO	24
3.6	ÍNDICE DE MILLER	26
3.7	DIFRAÇÃO DE RAIOS X	28
3.8	LEI DE BRAGG	31
3.9	O PROBLEMA DA FASE	36
3.10	SUBSTITUIÇÃO MOLECULAR	37
4	RESULTADOS E DISCUSSÃO	39
4.1	<i>CANAVALIA MARÍTIMA LECTIN</i> (CML) E <i>CANAVALIA GLADIATA LECTIN</i> (CGL)	39
4.2	SUBSTITUIÇÃO MOLECULAR	45
4.3	REFINAMENTO CRITALOGRÁFICO	48
4.4	DESCRIÇÕES DA ESTRUTURA	50
4.5	QUALIDADE DO MODELO	64
4.6	COMPARAÇÕES ENTRE OS SÍTIOS DE LIGAÇÃO	72
4.7	COMPARAÇÃO ENTRE A CML E A CGL	74
5	CONCLUSÃO	75
6	REFERENCIAS	76

1. INTRODUÇÃO

1.1 Lectinas

Lectinas são (glico) proteínas que possuem pelo menos um domínio não catalítico, denominado domínio lectínico, através do qual se liga de forma reversível e específica a mono ou oligossacarídeos. A primeira descrição sobre lectinas foi em 1888 publicada por Stillmark. Seu trabalho pioneiro foi um marco em biologia, porque foi o primeiro a propor relação entre a toxicidade de uma proteína encontrada na semente da mamona (*Ricinus communis*) e a ocorrência de um fator hemaglutinado protéico. Além disso, sua descoberta revolucionou a indústria bioquímica da época, pois ‘ricino’ foi à primeira proteína vegetal com atividades biológicas bem definidas que pode ser manipulada. Posteriormente, substâncias tóxicas similares foram identificadas em sementes de *Cróton tiglium*, *Abrus precatorius* e na casca de *Robinia pseudoacacia*.

Em meados do século XX, houve a constatação de que algumas hemaglutininas exibem clara preferência por eritrócitos de um determinado grupo sanguíneo humano, dentro do sistema ABO (Renkonen, 1948; Boyd and Reguera, 1949). A descoberta da especificidade do grupo sanguíneo foi o motivo que levou a introduzir o termo ‘lectina’(do verbo latim ‘legere’, que significa ‘para selecionar’). Embora, Howell (1936) tenha observado que o açúcar da cana-de-açúcar inibiu a atividade aglutinante da *Concanavalin A* (ConA), foi demonstrado apenas em 1952 que as propriedades aglutinantes de lectinas são baseadas em atividades específicas do açúcar vinculado (Watkins and Morgan, 1952).

Atualmente, lectinas abrangem um grupo muito grande de proteínas vegetais, desse modo, visto que as definições para as lectinas até então não estabeleciam uma

maneira de classificar tais proteínas de uma forma generalizada, algumas subdivisões foram criadas baseadas em aspectos estruturais. Peumans & Van Damme (1998) classificaram as lectinas em quatro grupos principais: merolectinas, hololectinas, quimerolectinas e superlectinas.

a) **Merolectinas** consiste de proteínas com um único domínio de ligação de açúcar (sítio de ligação), por definição são monovalentes e monoméricas, conseqüentemente, devido ao fato de possuírem um único sítio de ligação a carboidratos não apresentam atividade hemaglutinante ou capacidade de precipitar glicoconjugados. São consideradas, estruturalmente, as mais simples. Um exemplo comum de uma merolectina é a heveína, lectina presente no látex da seringueira, *Hevea brasiliensis*



(Figura 01).

Figura 01: Representação esquemática de merolectina (Peumans & Van Damme *et al.*, 1998).

b) **Hololectinas** são proteínas formadas por dois ou mais sítios de ligação de açúcar que sejam idênticos ou homólogos e tendem a ligar o mesmo tipo de açúcar. Hololectinas são por definição divalentes ou polivalentes (oligoméricas) e, conseqüentemente, aglutina células e/ou precipitam glicoconjugados. Apresentam-se, normalmente, como dímero ou tetrâmero. Uma grande quantidade de lectinas presentes

em plantas é classificada como hololectinas, dentre as quais podemos citar as lectinas de *Canavalia brasiliensis* (ConBr), *Artocarpus integrifolia* (Jacalina) e *Griffonia simplicifolia* (Figura 02).

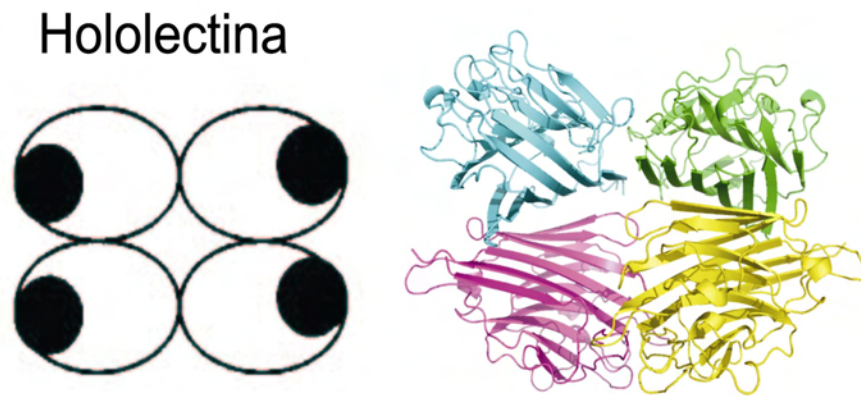


Figura 02: Representação esquemática de Hololectina (Peumans & Van Damme *et al.*, 1998).

c) **Quimerolectinas** são fusões de proteínas tendo um ou mais sítio de ligação de açúcar, entretanto, um sítio de ligação deve ser preenchido por um carboidrato com atividade biológica bem definida agindo independentemente dos outros sítios de ligação. Neste grupo geralmente são encontrados proteínas com dupla função, mas que são complementares entre si como, por exemplo, proteínas relacionadas com invasividade celular, as quais utilizam o domínio de reconhecimento a açúcar como ponto para ancoragem e subsequentemente age com o outro domínio para efetivar a entrada ou até mesmo a destruição de membranas. Dependendo do número de sítios de ligação, quimerolectinas podem se comportar como merolectinas ou hololectinas. Por exemplo, temos o caso das RIPs do tipo 2. Estas apresentam mais de um domínio de ligação a carboidrato comportando-se, também, como hololectinas (Figura 03).

Quimerolectina

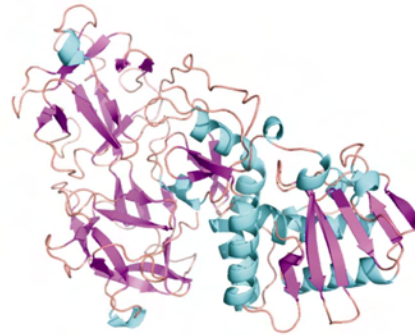
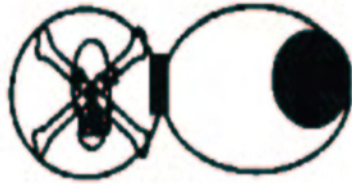


Figura 03: Representação esquemática de quimerolectina (Peumans & Van Damme *et al.*, 1998).

d) Superlectinas, como as hololectinas, são proteínas formadas por dois ou mais sítio de ligação de açúcar, entretanto, seu domínio catalítico pode reconhecer diferentes tipos de açúcar. São menos encontradas do que as hololectinas, porém novas estruturas para esse grupo de lectinas vêm sendo resolvidas como, por exemplo, a estrutura da banana-lectin que possui dois sítios distintos de reconhecimento a laminaribiose e Xyl-beta1, 3-Man-alpha-O-Methyl (Meagher *et al.*, 2005). Em alguns casos, superlectinas podem ser consideradas casos especiais de quimerolectinas (Figura 04).

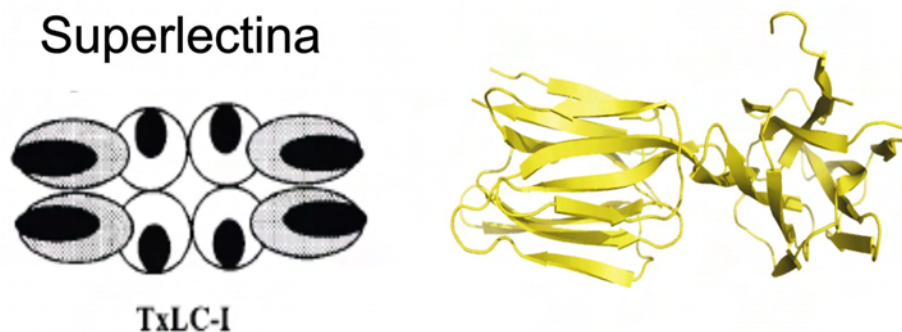


Figura 04: Representação esquemática de Superlectina (Peumans & Van Damme *et al.*, 1998).

e) Embora essa classificação tenha sido feita, inicialmente, em quatro grupos abrangendo significativamente inúmeras proteínas vegetais até então descobertas, algumas lectinas de plantas, hoje descobertas e estruturalmente resolvidas, não se enquadram especificamente em nenhuma destas classificações. Como exemplo podemos citar a lectina, recém descoberta, presente na semente de *Parkia platycephala*, homóloga à família das hidrolases. Esta lectina possui um domínio com capacidade de reagir enzimaticamente com o polissacarídeo composto de monômeros de *N*-acetyl-D-glucosamina (quitina) e, ao mesmo tempo, dentro deste mesmo domínio, possui outro sítio de reconhecimento a carboidrato (Cavada *et al.*, 2006) (figura 05).

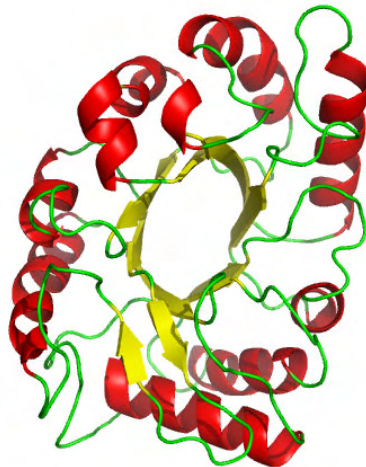


Figura 05: Lectina presente na semente de *Parkia platycephala* que apresenta o tipo de enovelamento Barril β . Essa lectina não se enquadra especificamente em nenhuma das classificações propostas por Peumans & Van Damme (Cavada *et al.*, 2006).

1.2 Famílias evolutivas

O avanço da biologia estrutural em conjunto com modernas técnicas de DNA recombinante deu um salto significativo na elucidação das mais diversas moléculas presentes em estruturas dos sistemas biológicos. Atualmente mais de 1000 estruturas de lectinas de vegetais encontram-se depositadas no Protein Data Bank. Muitas dessas lectinas possuem seus sítios de ligação a carboidratos bem definidos tendo um detalhamento estrutural a nível atômico (alta resolução).

Dentre todas essas lectinas estudadas até agora, observa-se uma grande semelhança em aspectos estruturais, como sequência primária, reconhecimento a carboidratos, e função biológica. Tais semelhanças ou discrepâncias são puramente reflexos do processo evolutivo que vem ocorrendo com essas proteínas a milhares de

anos. Dessa forma é possível classificar essas proteínas em famílias evolutivamente relacionadas, ou seja, que apresentem estrutura quaternária semelhante e compartilhem outras características que possam estar diretamente relacionadas com a função biológica desempenhada por cada uma.

Dentre as famílias evolutivamente relacionadas de lectinas de plantas, podemos destacar:

- a) Lectinas de leguminosas;
- b) Lectinas específicas à quitina e homólogas a heveína;
- c) Lectinas homólogas ao tipo RIP-2;
- d) Lectinas homólogas a Jacalina;
- e) Lectinas da família das Amarantaceas;
- f) Lectinas de floema de Curcubitaceae.

Essa classificação, feita em 1998 (Van Damme *et al.*, 1998), não é exaustiva, pois novas classificações surgem no decorrer das descobertas. Inclusive, sabe-se que outras lectinas já são consideradas pertencentes a grupos diversos como o das *Apiaceae*, *Araucariaceae*, *Celastraceae*, *Euphorbiaceae*, *Gramineae*, *Labiatae*, *Solanaceae*, etc.

De todos esses grupos de famílias evolutivamente relacionadas citado acima, o grupo mais estudado é o das lectinas de leguminosas. Inúmeras estruturas desse grupo foram obtidas nos mais diversos grupos de pesquisa, os quais conseguiram ao longo de décadas estabelecer padrões estruturais que definem com precisão quais características são comuns para todos exemplares até então investigados.

1.3 Lectinas de leguminosas

A história do estudo de lectinas de leguminosas começa em 1890 com o relatório intitulado **Power and Cambier** sobre o isolamento de uma lectina tóxica chamada ‘robin’ encontrada na casca de uma leguminosa conhecida como gafanhoto negro (*Robinia pseudoacacia*) (Kocourek, 1896). Em um trabalho posterior, Landsteiner e Raubitschek (1907) descobriram a presença de lectinas não tóxicas em diversas leguminosas, tais como: *Phaseolus vulgaris* (feijão), *Pisum sativum* (ervilha), *Lens culinaris* (lentilha) e *Vicia sativa*. Esse trabalho demonstrou pela primeira vez que lectinas não são necessariamente proteínas tóxicas. A partir desse feito, inúmeras pesquisas foram realizadas com essas lectinas. Finalmente em 1936, conseguiu-se purificar e cristalizar a primeira lectina vegetal (ConA) e posteriormente resolver sua estrutura tridimensional (Edelman *et al.*, 1972; Hardman and Ainsworth, 1972).

O termo ‘lectinas de leguminosas’ refere-se a um tipo particular de lectinas de plantas que são encontradas exclusivamente na família das *Leguminosae*. Porém convém ressaltar que nem todas as lectinas encontradas em leguminosas pertencem ao grupo de lectinas de leguminosas. Por exemplo, as lectinas encontradas nas sementes do *Abrus precatorius* são do grupo tipo 2 RIP. Por ser o grupo mais conhecido dentre a família de lectinas, inúmeras lectinas de leguminosas vem se tornando ferramentas biológicas indispensáveis na pesquisa biomédica, sendo amplamente utilizadas como proteínas bioativas.

As lectinas desse grupo normalmente possuem cerca de 220 a 250 resíduos de aminoácidos (cerca de 30 kDa) por monômero e são extremamente semelhantes quanto aos aspectos estruturais. Apresentam o mesmo tipo de enovelamento em relação a seus monômeros. Esses monômeros são compostos exclusivamente de sete fitas beta (**front**

face) e seis fitas beta (**back face**) interligadas por voltas e alças formando uma estrutura na forma achatada (figura 06). Esse tipo de enovelamento é conhecido como “ β -sanduíche”, por ser composto de fitas betas antiparalelas (cerca de 60%). A orientação antiparalela das fitas beta e a mudança entre uma face e outra, forma uma estrutura estável, explicando em parte a extrema resistência de lectinas de leguminosas à ação de muitas enzimas. Quatro **loops** localizados na parte superior da lectina (estrutura tridimensional) formam o sítio de ligação do monômero, onde é complexado o açúcar. Como quase não há estruturas helicoidais, todos os monômeros de lectinas de leguminosas pertencem à classe de proteínas beta (Richardson, 1981).

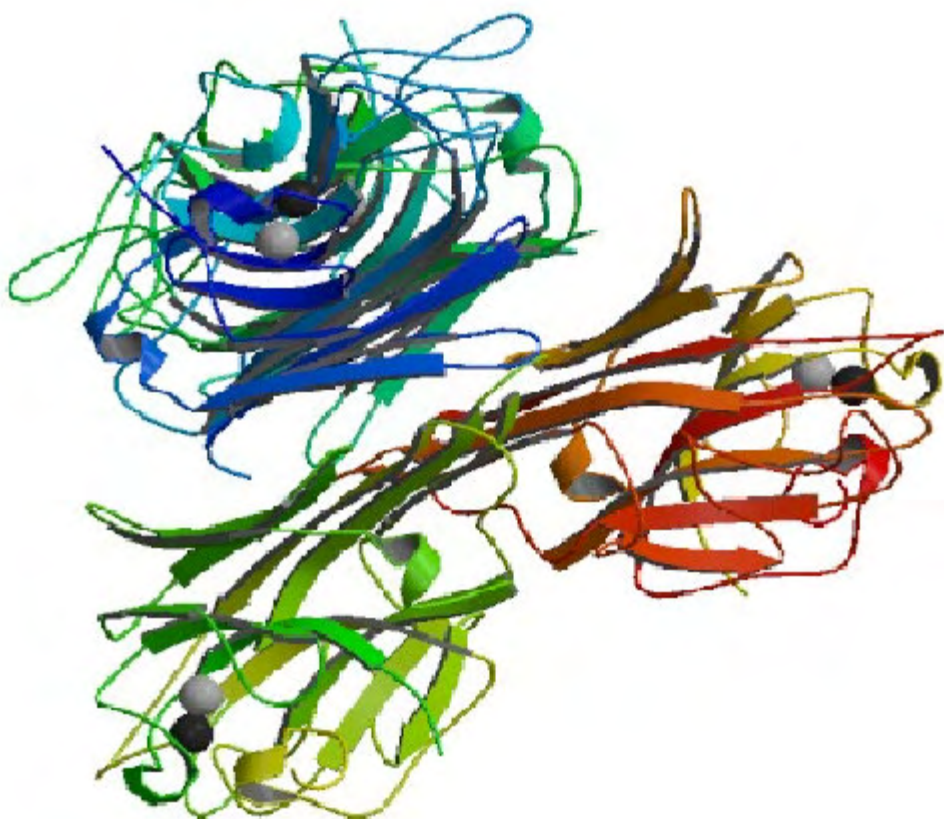


Figura 06: Estrutura da lectina de leguminosa CGL man 1-4 man (PDB: 2p2k).

Lectinas de leguminosas podem interagir tanto com açúcar simples (monossacarídeos) quanto açúcar complexo (oligossacarídeos) (Drickamer, 1995; Weis e Drickamer, 1996). Alguns resíduos de aminoácidos pertencentes a quatro **loops** localizados no topo da estrutura, formam o sítio de ligação do monômero e são responsáveis pelo reconhecimento de tais açúcares. Tomando como exemplo a lectina LoLI podemos ver que os resíduos (Gly²⁹, Ala³⁰, Glu³¹, Asp⁸¹, Gly⁹⁹ e Asn¹²⁵) formam uma rede com oito pontes de hidrogênio com um açúcar simples (figura 07). Além disso, um resíduo aromático (phe¹²³) cria uma interação hidrofóbica com o anel piranose do açúcar.

Todo monômero de lectinas de leguminosas possui uma cavidade hidrofóbica, descrita pela primeira vez para a ConA (Becker *et al.*, 1975). Essa concavidade conserva os resíduos hidrofóbicos presentes em cada monômero.

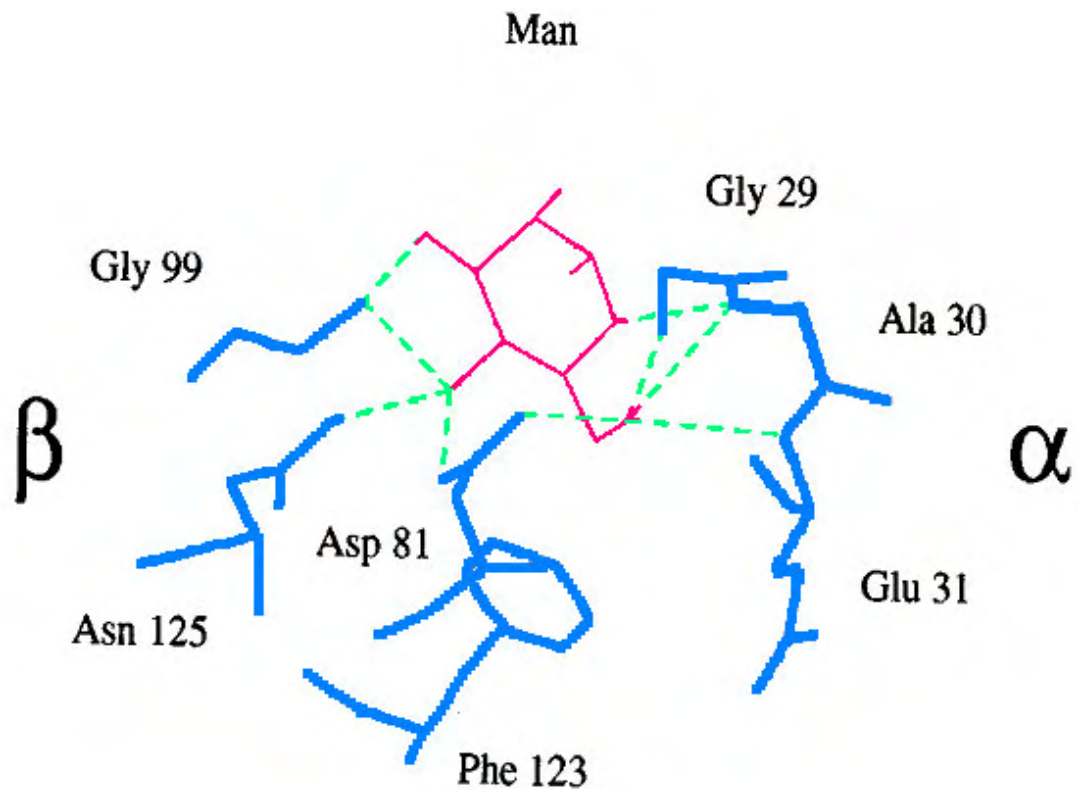


Figura 07: Redes de pontes de hidrogênio (presentes na lectina LoLI) formadas entre seis resíduos de aminoácidos e um açúcar simples. O resíduo aromático (phe¹²³) cria uma interação hidrofóbica com o anel piranose do açúcar.

Cada monômero geralmente possui dois ou mais sítios de ligação a íons como também podem ser dotados de um ou mais sítios de ligação a carboidratos. O sítio de ligação a íons é amplamente conservado nestas proteínas. Podemos observar que a capacidade de reconhecer o açúcar depende da presença dos íons cálcio (Ca^{++}) e de um outro, geralmente relatado como sendo o manganês (Mn^{++}). Esses metais estão localizados a uma distância de aproximadamente 4.5 Å um do outro, e são coordenados por cadeias laterais constituídas de oito aminoácidos e três moléculas de água.

Nesse grupo de proteínas, os íons metálicos são ligados reversivelmente aos monômeros, dessa forma, a remoção desses íons resultam em importantes mudanças conformacionais. Tomando como exemplo a ConA, o peptídeo Ala²⁰⁷-Asp²⁰⁸, para

participar efetivamente do reconhecimento ao seu carboidrato específico necessita estar na conformação *trans*, conformação essa somente obtida com a presença dos íons cálcio e manganês, caso contrario esse assume a forma *cis* (conformação pouco favorável). Assim, as lectinas de leguminosas podem assumir duas conformações quanto à presença dos íons cálcio e manganês. Na presença destes assumem a conformação “**locked**” e na ausência “**unlocked**”. Portanto, a conformação mais favorável do sítio de ligação a carboidratos das lectinas de leguminosas ocorre exclusivamente pela presença dos íons cálcio e manganês (Bouckaert *et al.*, 1999).

2. Objetivos

O presente trabalho tem como objetivos a cristalização e a resolução da estrutura tridimensional das lectinas *Canavalia gladiata* (CGL) e *Canavalia marítima* (CML) complexadas com o açúcar manose 1-6 manose. O estudo dessas lectinas visa proporcionar comparações com outras estruturas (CGL e CML complexadas ao açúcar manose 1-2 manose, manose 1-3 manose e manose 1-4 manose) na intenção de identificar possíveis diferenças ou similaridades no sítio ativo entre essas lectinas.

3. Materiais e Métodos

3.1 Cristalização de Macromoléculas Biológicas

Biocristalografia é a aplicação das técnicas de difração de raios X para cristais de macromoléculas biológicas. O uso dessa técnica, juntamente com estudos bioquímicos, tem contribuído para esclarecer a base molecular da atividade biológica dessas macromoléculas (Blundell & Johnson, 1976).

Um dos passos essenciais para determinar a estrutura tridimensional de uma macromolécula por técnicas de difração de raios X, consiste primeiramente na obtenção de monocristais de tal proteína que apresentem um padrão de difração adequado. Esse processo envolve algumas etapas importantes:

- i.** De uma forma geral, quanto mais pura estiver a proteína, maiores as chances de crescimento de cristais.
- ii.** A proteína é dissolvida em uma solução tampão contendo aditivos e agentes precipitantes como sulfato de amônio, cloreto de sódio e outros.
- iii.** A solução deve atingir um estado de supersaturação para que pequenos agregados de proteínas possam ser formados. Uma vez que os núcleos são formados, o crescimento dos cristais pode começar.

Cristalização é um processo de ordenação, no qual as moléculas assumem posições regulares no estado sólido. Um processo de cristalização pode ser considerado como um equilíbrio dinâmico entre as partículas no estado líquido e no estado sólido. Sendo assim, o cristal cresce quando o equilíbrio é deslocado no sentido de solidificação.

Para crescer cristais de qualquer composto, as moléculas ou íons têm de ser trazidos a um estado de supersaturação, um estado que é termodinamicamente instável, no qual pode se desenvolver numa fase amorfa ou cristalina quando este retornar ao equilíbrio. A Supersaturação pode ser alcançada pela evaporação lenta do solvente ou pela variação de alguns dos parâmetros como temperatura, pH e força iônica. Neste estado, a solução de proteína alcança a zona de nucleação, onde podem se formar os núcleos (primeiros agregados ordenados da proteína). Com a diminuição da solubilidade da proteína, estes núcleos atingem a zona metaestável, onde ocorrerá o crescimento do cristal (Figura 08) (Ducruix & Giegé, 1992).

O conhecimento das solubilidades macromoleculares é um pré-requisito para o controle das condições de cristalização. Entretanto, como as bases teóricas da solubilidade apresentam controversas, especialmente no que tange efeitos dos sais, os dados sobre a solubilidade quase sempre se originam de determinações experimentais.

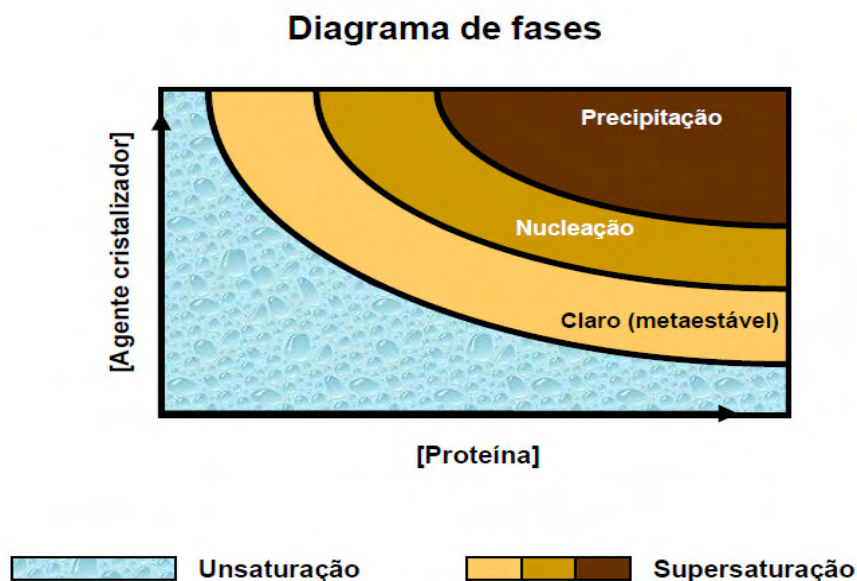


Figura 08: Representação esquemática de um diagrama de solubilidade para uma macromolécula quando a concentração do agente cristalizante é alterada.

A curva de solubilidade divide as regiões de hiposaturação e de supersaturação. Na região de hiposaturação a macromolécula não será cristalizada. Logo acima se encontra a região de supersaturação, que é dividida em três zonas: zona metaestável, zona de nucleação e zona de precipitação. Quando a supersaturação é alcançada rapidamente, a proteína em solução precipita. Porém, quando a supersaturação é alcançada lentamente, a solução de proteína alcança a zona de nucleação, onde são formados os núcleos e posteriormente esses núcleos atingem a zona metaestável, onde ocorrerá o crescimento do cristal. Por outro lado, pode ocorrer da proteína alcançar a zona de supersaturação rapidamente, e conseqüentemente precipitar, mas logo após, devido à diminuição de solubilidade, parte da proteína presente na gota atinge a zona de nucleação e posteriormente a zona metaestável.

A cristalização de macromoléculas é composta por três passos distintos: nucleação (isto é, a formação dos primeiros agregados ordenados), crescimento e

término do crescimento. Devido ao fato que proteínas e ácidos nucléicos necessitem de pH e força iônica definidas para a sua estabilidade e função, cristais de macromoléculas biológicas têm de ser crescidos a partir de soluções aquosas complexas. As condições para se atingir a nucleação, são algumas vezes difíceis de repetir, e desta forma os procedimentos de sementeação com um material cristalino pré-formado se faz necessário. Em muitos casos este é o único método para se obter resultados reprodutíveis. Após a fase de nucleação, passamos para o segundo passo, a fase de crescimento, onde os núcleos de cristais formados na zona de nucleação, “migram” para a zona metaestável, responsável pelo crescimento do cristal. A nucleação requer um estado de supersaturação maior do que aquele da fase de crescimento. Por fim, temos o término do crescimento de um cristal de macromolécula, onde o cristal já possui em muitos dos casos, dependendo da pureza do cristal, faces bem definidas.

3.2 Métodos de cristalização de macromoléculas biológicas

Dentre os vários métodos de cristalização, o mais utilizado é a difusão de vapor, sendo possíveis pelo menos três variações, sendo elas: a gota suspensa (*hanging drop*), gota sentada (*sitting drop*) e a gota sanduíche (*sandwich drop*) (figura 09).

O princípio destes métodos é o processo de equilíbrio entre duas soluções através da fase de vapor num meio fechado. A solução menos concentrada perderá seu solvente volátil até que os potenciais químicos das duas sejam iguais.

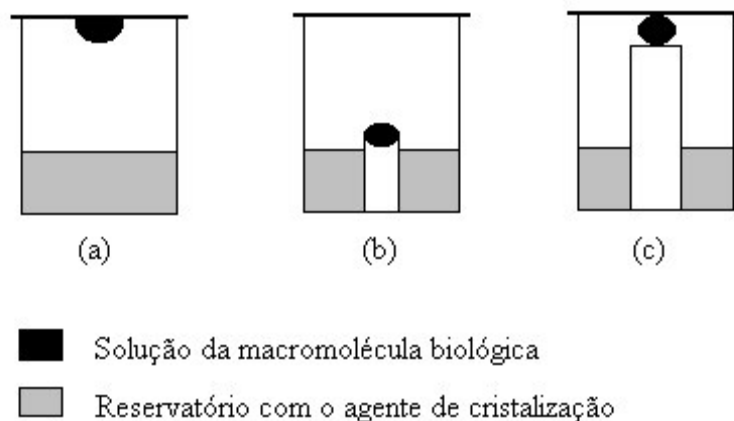


Figura 09: Representação esquemática de (a) gota suspensa (**hanging drop**), (b) gota sentada (**sitting drop**) e (c) gota sanduíche (**sandwich drop**).

Para cristalização das proteínas envolvidas neste trabalho, foi utilizado o método de difusão de vapor, sistema hanging drop (gota suspensa) (figura 09 (a)) através do método da matriz esparsa, desenvolvido pela Dra. J. Jancarik, que consiste de 96 soluções para cristalização (Jancarik; Kim, 1991).

O aparato necessário para desenvolver a etapa de cristalização de macromoléculas biológicas é: placas *Linbro*, graxa a vácuo, lamínulas siliconizadas, soluções do método da matriz esparsa e a macromolécula biológica a ser cristalizada.

Cada placa de cristalização é constituída por 24 reservatórios. Após distribuir cada solução do método da matriz esparsa em seus respectivos reservatórios, foi colocada uma camada de graxa a vácuo nas bordas dos reservatórios. Em cada lamínula siliconizada (22x22mm), foi pipetado 1µL da solução de proteína complexada com o substrato e sobre essa, foi acrescentado 1µL da solução presente em cada reservatório (líquido mãe). Os reservatórios (contendo 500µL cada) foram fechados com suas respectivas lamínulas, proporcionando assim, o sistema fechado necessário para dar início ao processo de difusão de vapor (figura 10). A proporção (1µL : 1µL) foi

utilizada para que a concentração da mistura presente na gota ficasse 50% menor que a concentração da solução precipitante do reservatório.

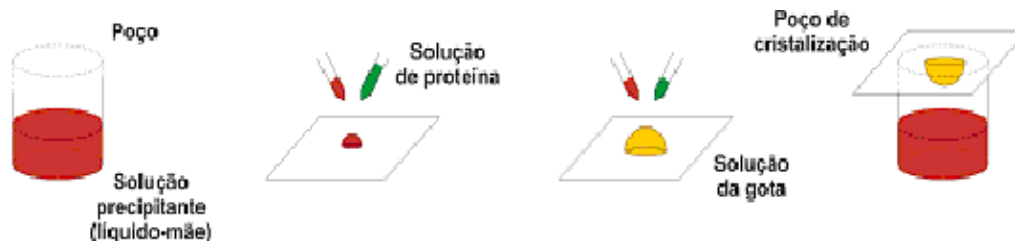


Figura 10: Método da gota suspensa (**hanging drop**) para a cristalização de proteínas. O equilíbrio é alcançado pela difusão de vapor entre a gota e a solução precipitante.

Assim, cada gota contendo a macromolécula biológica a ser cristalizada em tampão com o agente de cristalização e aditivos é equilibrada contra um reservatório contendo a solução do agente de cristalização em uma concentração 50% mais alta que a da gota. O equilíbrio prossegue por meio da difusão das espécies voláteis (água e solventes orgânicos) até que a pressão de vapor na gota se iguale aquela do reservatório. Se o equilíbrio ocorre por meio de troca de água (da gota para o reservatório), isto leva a uma diminuição do volume da gota, e conseqüentemente a um aumento na concentração de todos os constituintes da gota de cristalização.

Os cristais de macromoléculas são mantidos por interações como pontes de hidrogênio, pontes salinas, interações de Van der Waals e interações entre dipolos nas regiões de contato (McPherson, 1982; Ducruix; Giege, 1992). Uma grande parte da superfície dessas macromoléculas interage com o solvente, que preenche o volume cristalino não ocupado por elas (Tsay, *et al*, 1999). Assim, cristais protéicos são compostos de duas fases diferentes (sólida e líquida), por isso são frágeis e sensíveis ao ambiente (temperatura, condições da solução, choques, etc).

3.3 Cristais

O estudo dos cristais começou em 1669 quando Nicolaus Steno propôs que, durante o crescimento dos cristais, os ângulos entre as faces permaneciam constantes. A regularidade interna dos cristais foi sugerida, mas, somente em 1912, ela foi provada por Max von Laue, Walter Friedrich e Paul Knipping na primeira experiência de difração de raios X.

Um cristal é um arranjo periódico de átomos. Um cristal ideal é constituído pela repetição infinita de uma mesma estrutura elementar (sem defeitos e impurezas). Nos cristais mais simples, a unidade estrutural é representada por um único átomo. A estrutura de todos os cristais pode ser descrita em termos de uma rede com um grupo de átomos e íons ligados a cada ponto da rede. Esse grupo denomina-se base; ele repete-se infinitamente por todo espaço do cristal, formando assim a estrutura cristalina.

3.4 Estrutura cristalina

O fato fundamental e mais importante relativo a uma substância cristalina é que os elementos de que são compostos estão dispostos de maneira ordenada. A geometria do arranjo das unidades que definem uma determinada estrutura cristalina pode ser imaginada como sendo pontos que se repetem ao longo de direções e distâncias definidas, de forma que todos os pontos tenham vizinhanças idênticas.

Se um ponto da rede for escolhido ao acaso como sendo a origem da rede cristalina, a posição de qualquer outro ponto será definida pelo vetor $\mathbf{T}_{n_1, n_2, n_3} = n_1 \mathbf{a} + n_2 \mathbf{b} + n_3 \mathbf{c}$, onde \mathbf{a} , \mathbf{b} e \mathbf{c} são os vetores de módulo a , b e c ao longo das 3 direções não-

coplanares e n_1 , n_2 e n_3 são números inteiros¹. Os vetores **a**, **b** e **c**, bem como os ângulos α , β e γ entre eles, definem um paralelepípedo chamado de cela unitária. As linhas nas direções de **a**, **b** e **c** são chamadas de eixos x , y e z , respectivamente.

A escolha da cela unitária pode ser feita de várias maneiras. Celas unitárias contendo apenas um ponto da rede são chamadas primitivas, e aquelas contendo mais de um ponto são chamadas múltiplas ou centradas (figura 11).

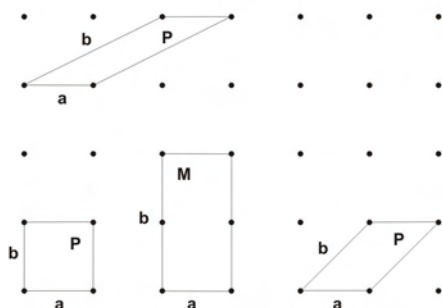


Figura 11: As células marcadas com a letra **P** são primitivas, e a célula marcada com a letra **M** é denominada múltipla ou centrada.

O conteúdo de uma cela unitária varia de cristal para cristal e pode conter desde poucos íons até algumas dezenas de moléculas com milhares de átomos. Além disso, pode-se ver que a estrutura de um cristal é obtida pela repetição da cela unitária em cada ponto da rede, ou seja, pela aplicação da operação de translação $\mathbf{T}_{n_1, n_2, n_3} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}$ a uma cela unitária (figura 12).

¹ Em alguns casos, dependendo da escolha dos vetores **a**, **b** e **c**, os índices n_1 , n_2 e n_3 podem assumir valores racionais. Um exemplo disso pode ser visto em cristais que apresentam células unitárias múltiplas ou centradas.

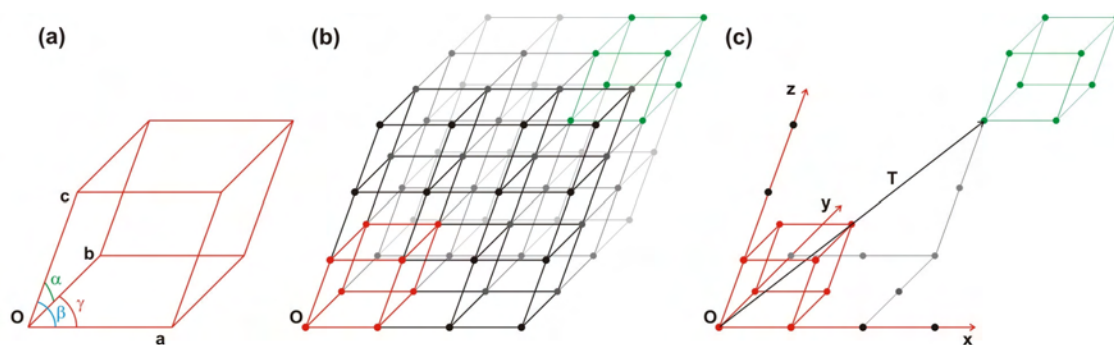


Figura 12: Representação de uma cela unitária geral. (a) Cela unitária com eixos **a**, **b** e **c** e ângulos α , β e γ situada na origem **O**. (b) A rede cristalina é obtida com o arranjo tridimensional das células unitárias. (c) Qualquer cela unitária do cristal está relacionada com a cela da origem por uma operação de translação **T**.

O paralelepípedo formado pelos eixos **a**, **b** e **c** denomina-se célula primitiva (ou cela unitária) e seu volume pode ser calculado através da fórmula:

$$V_c = (a \times b) \cdot c \quad (1)$$

Uma cela unitária é usada para especificar um dado arranjo de pontos em um retículo cristalino, em um conjunto de eixos, com a origem ajustada em um dos pontos do retículo. Para cada retículo espacial teremos um arranjo conveniente de coordenadas. Existem 7 sistemas² de coordenadas tridimensionais usadas para descrever um cristal e serve como base para sua classificação (tabela 01).

² Estes sistemas impõem restrições nas redes cristalinas exceto para o sistema triclínico que é a identidade.

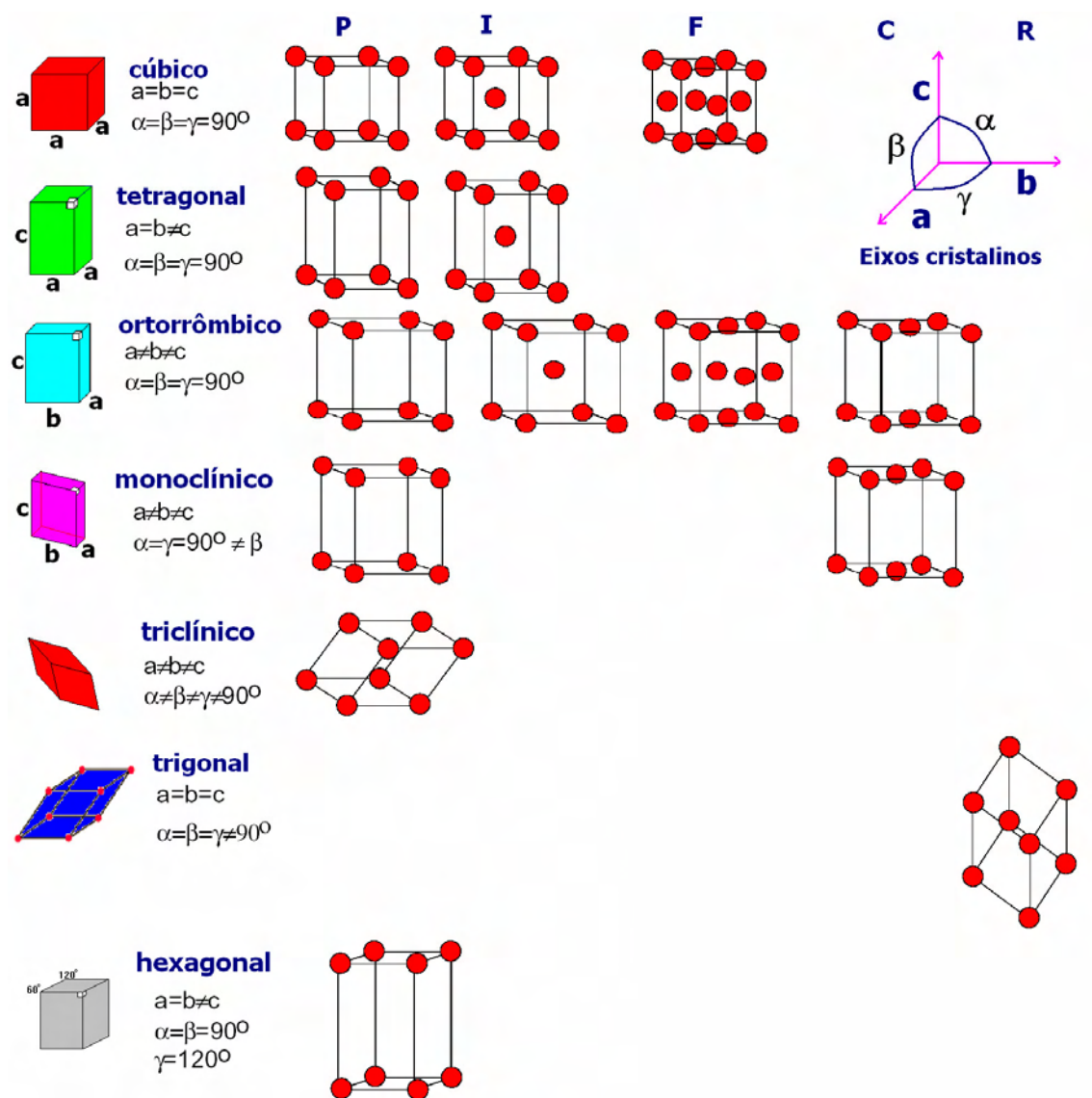
Tabela 01: Sistemas Cristalinos

Sistema	Eixos	Ângulos	Números de retículos
Cúbico	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$	3
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$	2
Ortorrômico	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$	4
Monoclínico	$a \neq b \neq c$	$\alpha = \gamma = 90^\circ, \beta$	2
Trigonal	$a = b = c$	$\alpha = \beta = \gamma < 120^\circ$	1
Hexagonal	$a = b \neq c$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$	1
Triclínico	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$	1

Quando a cela unitária possui unicamente pontos reticulares nos vértices do paralelepípedo que a compõem, é chamada primitiva, e denotada pela letra **P** (figura 11) (exceção feita para o sistema romboédrico, em que é denotado pela letra **R**). Aos retículos **P**, juntam-se os sete sistemas não-primitivos, que possuem a mesma simetria, mas comportam diversos pontos na cela unitária: são os retículos de corpo centrado, que apresentam um ponto no centro do paralelepípedo que compõem a cela unitária e são denominados pela letra **I**; os retículos de faces centradas, que apresentam um ponto no centro de cada face e são denominados pela letra **F**; e os retículos de bases centradas **A**, **B** e **C**, estes últimos, apresentam um ponto no centro de suas bases.

Assim, essas estruturas elementares são denominadas redes de Bravais³, em homenagem a Auguste Bravais que em 1848, a partir dos estudos publicados em 1842 por Moritz Ludwig Frankenheim (1801-1869), demonstrou geometricamente que num espaço tridimensional só poderiam existir 14 configurações básicas (figura 13).

³ Possíveis variáveis do estado primitivo levando em consideração os conjuntos de restrições de cada sistema.



3.5 Espaço recíproco

Para cada retículo cristalino é possível construir um retículo recíproco, assim chamado porque muitas das suas propriedades são recíprocas às propriedades do retículo cristalino. Considerando um retículo cristalino que possua uma cela unitária

definida pelos vetores **a**, **b**, **c** definimos uma cela unitária do retículo recíproco pelos vetores, **a***, **b*** e **c*** dados por:

$$\mathbf{a}^* = \frac{1}{V} (\mathbf{b} \times \mathbf{c}) = \frac{1}{V} bc \sin \alpha \quad (2)$$

$$\mathbf{b}^* = \frac{1}{V} (\mathbf{c} \times \mathbf{a}) = \frac{1}{V} ac \sin \beta \quad (3)$$

$$\mathbf{c}^* = \frac{1}{V} (\mathbf{a} \times \mathbf{b}) = \frac{1}{V} ab \sin \gamma \quad (4)$$

onde V é o volume da cela unitária.

$$V = abc \sqrt{1 + 2 \cos \alpha \cos \beta \cos \gamma - \cos^2 \alpha - \cos^2 \beta - \cos^2 \gamma} = \frac{1}{V^*} \quad (5)$$

A posição dos nós, pontos onde os três planos que formam a rede direta se cruzam, é representada por um vetor $\mathbf{H} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$, onde (u, v, w) são inteiros que definem as coordenadas do ponto na extremidade do vetor \mathbf{H} . A posição dos nós da rede recíproca será definida por um vetor $\mathbf{H}_{hkl}^* = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$, onde (h, k, l) são inteiros. Colocando os eixos da rede recíproca, normais aos planos (definidos por pares de eixos) da rede direta ($\mathbf{a}^* \wedge \mathbf{b}$ e \mathbf{c} , $\mathbf{b}^* \wedge \mathbf{a}$ e \mathbf{c} , $\mathbf{c}^* \wedge \mathbf{a}$ e \mathbf{b}) pode-se definir a rede recíproca diretamente das equações:

$$\mathbf{a}^* \cdot \mathbf{a} = 1 \quad \mathbf{a}^* \cdot \mathbf{b} = 0 \quad \mathbf{a}^* \cdot \mathbf{c} = 0 \quad (06)$$

$$\mathbf{b}^* \cdot \mathbf{a} = 0 \quad \mathbf{b}^* \cdot \mathbf{b} = 1 \quad \mathbf{b}^* \cdot \mathbf{c} = 0 \quad (07)$$

$$\mathbf{c}^* \cdot \mathbf{a} = 0 \quad \mathbf{c}^* \cdot \mathbf{b} = 0 \quad \mathbf{c}^* \cdot \mathbf{c} = 1 \quad (08)$$

onde **a**, **b** e **c** são vetores da rede cristalina (direta) e **a***, **b*** e **c*** são vetores da rede recíproca.

No caso, onde os eixos das duas redes são perpendiculares entre si, os eixos da rede recíproca terão comprimentos que são o recíproco dos comprimentos dos eixos da rede cristalina (direta). O espaçamento d_{hkl} entre os planos (hkl) da rede direta, em qualquer sistema cristalino, será o recíproco do comprimento do vetor \mathbf{H}_{hkl}^* .

$$\mathbf{H}_{hkl}^* = \frac{1}{d_{hkl}} \quad (9)$$

3.6 Índices de Miller

Em toda cela unitária de vetores **a**, **b** e **c** e ângulos α , β e γ localizada na origem **O** da rede cristalina, existem infinitos planos que a intercepta. Suponha-se, então, que um único plano, cortando todo o cristal, intercepte os vetores **a**, **b** e **c** nos pontos a/h , b/k e c/l , respectivamente, sendo h , k e l inteiros. Os números h , k e l são chamados de índices de Miller e são obtidos tomando os três menores inteiros que satisfaz a relação $(1/x \ 1/y \ 1/z)$. Por exemplo: Supondo um paralelepípedo de arestas 1,2,3 com um dos vértices na origem. Ou seja, os planos que interceptam os eixos x,y,z são: (1,0,0),(0,2,0) e (0,0,3) então a interseção é (1 2 3) invertendo os números temos $1 \ 1/2 \ 1/3$, os menores inteiros que satisfazem essa relação são os números (6 3 2) os quais são chamados de índices de Miller.

Em alguns casos, podem existir planos que não podem ser expressos como múltiplos de números inteiros multiplicados por um fator. Entretanto devido à periodicidade dos cristais somente os planos com índices como números racionais é que

são considerados para fins de difração, pois os planos irracionais não constituem um conjunto de planos idênticos, paralelos e eqüidistantes.

Para todo plano determinado por índice de Miller, pode-se estabelecer correlação com o espaço real dado pela formula:

$$hx+ky+lz= 1 \quad (10)$$

Por exemplo, os pontos (1,-2,0) e (3,-5,-2) determinados no espaço real, estão contidos no plano (21 10 6).

Para o plano apresentado abaixo (figura 14), os índices de Miller são (hkl) , se os inteiros h , k e l não tiverem divisores comuns. Também para os índices de Miller segue-se a convenção de colocar o sinal “-“ sobre os índices negativos; por exemplo, $(hk\bar{l})$ indica intersecção negativa ao longo do eixo c . Se um dado plano é paralelo a um dos eixos fundamentais, então não o intersecta e o valor do índice de Miller correspondente àquela direção é igual à zero. Por exemplo, a família de planos paralelos ao plano definido pelos vetores fundamentais **a** e **b** tem índices de Miller (001).

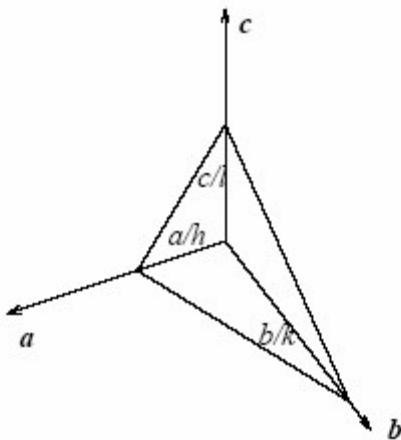


Figura 14: Planos cristalinos com índices (hkl) .

Se uma sequência de planos paralelos ao plano inicial é criada, mantendo a mesma distância d_{hkl} entre eles, uma família de planos (hkl) atravessará todos os pontos da rede cristalina (figura 15).

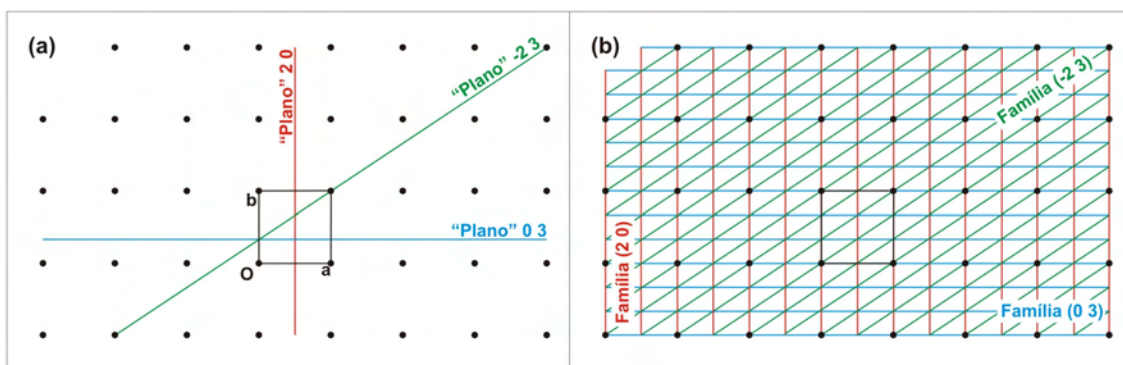


Figura 15: Uma rede cristalina bidimensional ilustrando os índices de Miller para certo conjunto de “planos”. (a) Rede cristalina com uma cela unitária situada na origem O e três “planos” cruzando os eixos a e b. (b) As três famílias de “planos” são obtidas com a repetição dos “planos” originais.

3.7 Difração de raios X

Os raios X foram descobertos em 8 de novembro de 1895 pelo físico alemão Wilhelm Conrad Roentgen (1845-1923), quando realizava experimentos com os raios catódicos e observou certa luminosidade partindo de uma tela recoberta com platinocianeto de bário que estava próxima ao tubo de vácuo. Após esse fato, diversos trabalhos foram publicados por vários autores sobre esses raios misteriosos, até então desconhecidos. Em 1901, Roentgen recebeu o prêmio Nobel de Física pela descoberta de notáveis raios que subsequentemente receberam seu nome.

Os raios X compreendem a faixa do espectro eletromagnético entre a radiação Ultra Violeta e a radiação Gama. Seu comprimento de onda varia entre 0.1 a 10\AA (figura 16). Eles apresentam propriedades típicas de ondas como polarização,

interferência e difração, da mesma forma que a luz e todas as outras radiações eletromagnéticas.

Os raios X são geralmente produzidos pela desaceleração rápida de elétrons transferindo energia de movimento em radiação. Desse modo podem ser gerados em tubos de raios X, onde se aplica uma diferença de potencial em um filamento de tungstênio (cátodo) que emite elétrons. Esses elétrons chocam contra um alvo metálico (ânodo), geralmente constituído de cobre ou molibdênio, gerando assim os raios X. Por esse método são produzidos dois tipos de raios X: Os que constituem o espectro contínuo que resulta da desaceleração do elétron durante a penetração no ânodo e os raios X característico do material do ânodo. Também podemos obter raios X através de isótopos radioativos como o ^{55}Fe e pela radiação Síncrotron.

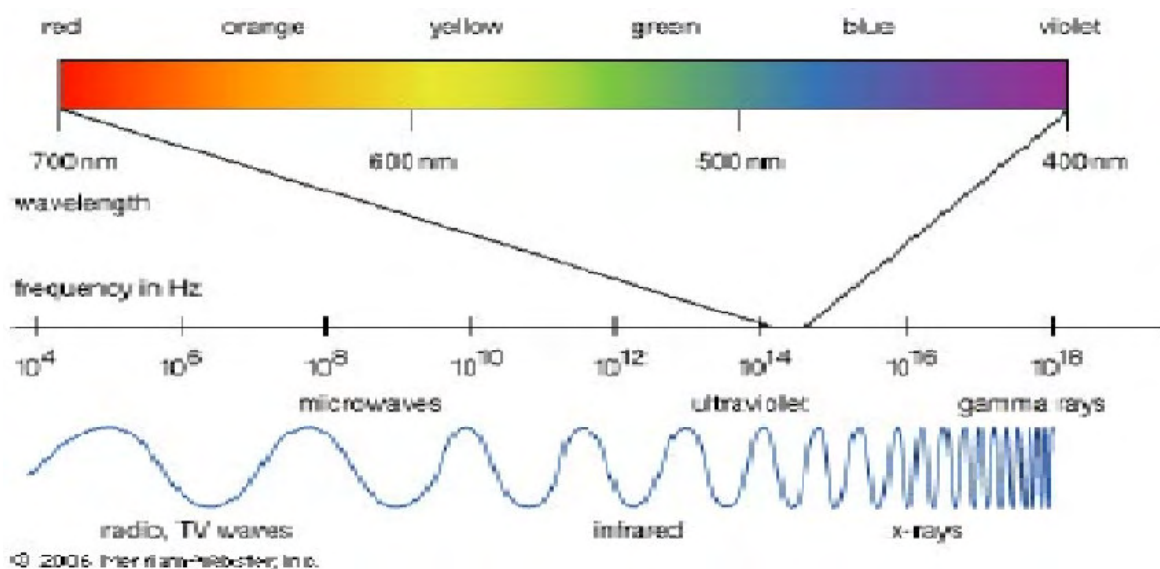


Figura 16: Espectro eletromagnético.

Em 1912 o físico alemão von Laue, sugeriu que se os átomos apresentam uma estrutura cristalina e são organizados de forma a apresentarem periodicidade ao longo do espaço, e que se os raios X fossem ondas eletromagnéticas com comprimento de onda menor que os espaços interatômicos, então os núcleos atômicos que concentram a massa dos átomos poderiam difratar os raios X, formando franjas de difração. Em seu experimento, Laue fez passar um feixe de raios X por uma amostra monocristalina, e pôs um filme fotográfico após a amostra; após revelar o filme, este apresentava pontos sensibilizados pelos raios X difratados.

Quando um feixe de raios X atravessa um cristal, ocorre uma interação entre o feixe e os elétrons dos átomos no cristal. Os elétrons passam a espalhar radiação (fótons de raios X) para todas as direções, num processo denominado espalhamento. A energia de um fóton de raios X é relacionada com seu comprimento de onda λ por $E = h\nu = hc/\lambda$, onde h é a constante de Plank, $h = 6,63 \cdot 10^{-27}$ erg.s = $6,62 \cdot 10^{-34}$ J.s. Em unidades mais usadas em laboratório: $\lambda (\text{\AA}) = 12,4/E(\text{keV})$.

Considerando a dualidade da luz, que pode ser tanto onda como partícula, podemos classificar duas formas de espalhamento: o espalhamento Compton, que leva em conta a natureza corpuscular da radiação e o espalhamento Thomson, que considera a radiação incidente como sendo uma onda eletromagnética plana e polarizada.

O espalhamento Compton ocorre quando raios X incidem sobre elétrons livres ou fracamente ligados e pode ser entendido a partir da teoria quântica, a partir da conservação do momento e da energia de colisão do fóton com o elétron. Assim temos que a radiação espalhada possui um comprimento de onda maior que o da radiação incidente (espalhamento inelástico) devido à transferência de energia do fóton para o elétron.

No espalhamento Thomson o campo elétrico oscilante associado ao feixe de raios X que incide sobre um elétron, obriga esse elétron a oscilar em torno da sua posição de equilíbrio. Sabemos que toda partícula carregada emite radiação, assim o elétron, submetido a um campo elétrico oscilante, emite uma onda eletromagnética, que possui o mesmo comprimento de onda da radiação incidente (espalhamento elástico).

3.8 Lei de Bragg

Em 1913, o físico inglês Sir W.H. Bragg e seu filho Sir W.L. Bragg apresentaram uma explicação para os feixes difratados por um cristal. Quando um feixe de raios X incide sobre um cristal, os planos de átomos deste cristal funcionarão como espelhos e farão com que o feixe seja refletido. Esta reflexão ocorre porque o espaçamento entre os planos atômicos de um cristal é da ordem do comprimento de onda do raio X. Para fundamentar matematicamente sua teoria, Bragg partiu das idéias de Laue.

Segundo a equação de Laue:

$$\mathbf{S} = \frac{\mathbf{S}' - \mathbf{S}}{\lambda} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \quad (11)$$

Para que ocorra difração em um cristal, o vetor \mathbf{S} tem que estar situado sobre um ponto do retículo recíproco, onde hkl são inteiros.

Fazendo o produto escalar entre a equação de Laue e os vetores **a**, **b** e **c**, temos:

$$\mathbf{a} \cdot \mathbf{S} = \mathbf{a} \cdot (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) = h \quad (12)$$

$$\mathbf{b} \cdot \mathbf{S} = \mathbf{b} \cdot (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) = k \quad (13)$$

$$\mathbf{c} \cdot \mathbf{S} = \mathbf{c} \cdot (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) = l \quad (14)$$

Quando essas três últimas equações são satisfeitas, um feixe de raios X difratado será produzido. Entretanto, o vetor **S** deve ser perpendicular ao plano *hkl* e θ o ângulo formado entre **s** (ou **s**₀) e os planos *hkl*. Assim:

$$\frac{2\sin\theta}{\lambda} = \left| \frac{s - s_0}{\lambda} \right| = |\mathbf{H}| = \frac{1}{d_{hkl}} \quad (15)$$

Assim, se a diferença de caminho entre os feixes for um múltiplo inteiro do comprimento de onda da radiação usada no experimento, então as ondas que representam cada um dos feixes “refletidos” irão se combinar com um máximo de interferência construtiva. Essa afirmação é conhecida como a Lei de Bragg e a equação que a quantifica é a seguinte:

$$n\lambda = 2d_{hkl}\sin\theta \quad (16)$$

onde *n* é um número inteiro, *d*_{*hkl*} é a distância entre os planos de Miller e λ o comprimento de onda incidente no cristal (figura 17). Como os raios X penetram profundamente nos cristais, um grande número de planos refletirá os feixes incidentes.

Porém, somente aquelas ondas refletidas por planos que tenham múltiplos inteiros do comprimento de onda incidente, irão interferir construtivamente enquanto as demais interferirão destrutivamente.

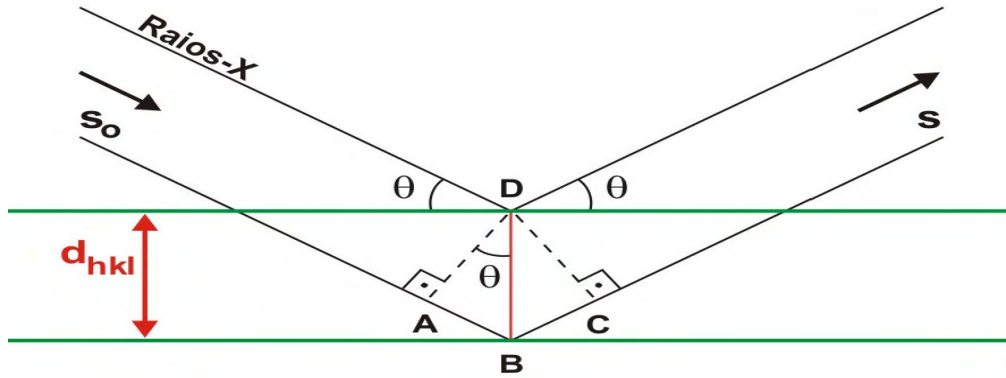


Figura 17: Esquema da lei de Bragg.

Toda propriedade física do cristal, tais como, concentração de cargas, densidade numérica dos elétrons, densidade dos momentos magnéticos e a densidade de massa, permanecem invariante sob qualquer translação do tipo $\mathbf{T}_{n_1, n_2, n_3} = n_1 \mathbf{a} + n_2 \mathbf{b} + n_3 \mathbf{c}$, sendo n_1 , n_2 e n_3 números inteiros e \mathbf{a} , \mathbf{b} e \mathbf{c} os eixos cristalinos. Assim, como a densidade numérica dos elétrons $n(\mathbf{r})$ é invariante, pode ser escrita sendo uma função periódica de \mathbf{r} com períodos \mathbf{a} , \mathbf{b} e \mathbf{c} nas direções dos eixos cristalinos, tendo a seguinte forma:

$$n(\mathbf{r}-\mathbf{T})=n(\mathbf{r}) \quad (17)$$

Sendo a densidade numérica dos elétrons uma função periódica pode-se representá-la como uma série de Fourier. Fazendo a análise de Fourier para funções periódicas em três dimensões encontramos um conjunto de vetores \mathbf{H} dado por:

$$n(\mathbf{r}) = \sum_H n_H \exp(i\mathbf{H} \cdot \mathbf{r}) \quad (18)$$

O conjunto de vetores \mathbf{H} da série de Fourier são os vetores da rede recíproca, valendo lembrar que toda estrutura cristalina possuem a rede cristalina (espaço real) e a rede recíproca (espaço imaginário). Os vetores da rede cristalina possuem dimensão de comprimento, enquanto que os vetores da rede recíproca possuem dimensão do inverso de seu comprimento. A rede cristalina é o espaço real do cristal; a rede recíproca é uma rede no espaço de Fourier associado. As duas redes podem ser relacionadas através da construção de vetores que definem os eixos da rede recíproca:

$$A = 2\pi \frac{b \times c}{a \bullet b \times c} \quad (19)$$

$$B = 2\pi \frac{c \times a}{a \bullet b \times c} \quad (20)$$

$$C = 2\pi \frac{a \times b}{a \bullet b \times c} \quad (21)$$

O conjunto de vetores \mathbf{H} da rede recíproca determina as possíveis reflexões dos raios X, sendo possível obter outra forma da expressão da lei de Bragg:

$$2\mathbf{k} \cdot \mathbf{H} + \mathbf{H}^2 = 0 \quad (22)$$

Este é o resultado central da teoria do espalhamento elástico numa rede periódica, resultado idêntico ocorrerá na teoria eletrônica sobre a estrutura das bandas de energia dos cristais (Kittel, 1978).

O conjunto de vetores \mathbf{H} da rede recíproca determina as possíveis reflexões dos raios X, assim, é possível introduzir os componentes de Fourier de $n(\mathbf{r})$ na equação da amplitude de onda espalhada que é dada por:

$$F = \int dV n(\mathbf{r}) \exp(-i\Delta\mathbf{k} \cdot \mathbf{r}) \quad (24)$$

onde \mathbf{r} é a posição da densidade de cargas espalhadas. Escrevendo os valores de $n(\mathbf{r})$ de acordo com a equação (18), temos:

$$F = \sum \int dV n_H \exp[i(H - \Delta\mathbf{k} \cdot \mathbf{r})] \quad (25)$$

Para obter o fator de estrutura F_H definido por uma integral sobre uma cela unitária com $\mathbf{r} = 0$ em um dos vértices, a condição $\Delta\mathbf{k} = \mathbf{H}$ tem que ser satisfeita, assim:

$$F_H = \int_{cel} dV n(\mathbf{r}) \exp(-i\mathbf{H} \cdot \mathbf{r}) = NF(hkl) \quad (26)$$

Reescrevendo o fator de estrutura F em função da densidade eletrônica $\rho(\mathbf{r})$, temos:

$$F(S) = \int_V \rho(\mathbf{r}) \exp 2\pi i(\mathbf{r} \cdot \mathbf{s}) d\mathbf{r} \quad (27)$$

onde $\Delta k = \mathbf{S} = \mathbf{s} - \mathbf{s}_0$ (direção da radiação espalhada menos o feixe incidente).

Multiplicando ambos os lados por $(\exp - 2\pi(\mathbf{r}' \bullet \mathbf{S}))$ e integrando sobre o volume recíproco ($V^* = 1/V$), temos:

$$\rho(\mathbf{r}) = \int_{V^*} F(\mathbf{S}) 2\pi(\mathbf{r}_i \bullet \mathbf{S}) dV^* \quad (28)$$

Essa integral pode ser substituída por uma somatória, pois $F(\mathbf{S})$ não é contínua e somente os pontos pertencente ao reticulado recíproco são diferentes de zero, assim:

$$\rho(x,y,z) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) \exp i\alpha(hkl) \exp[-2\pi i(hx + ky + lz)] \quad (29)$$

onde $|F(hkl)|$ é a amplitude do fator de estrutura das reflexões (hkl) , x_j , y_j e z_j são as coordenadas fracionárias do j ésimo átomo na cela unitária, h , k e l são os índices de reflexão, e $\alpha(hkl)$ é o ângulo da fase.

3.9 O Problema da Fase

Através da equação (28) fica explícita a necessidade de se conhecer o módulo e a fase dos fatores de estrutura, entretanto, pode-se obter somente o módulo do fator de estrutura a partir da difração de raios X; sendo toda informação sobre a fase perdida durante o processo. Como $I(hkl) \propto |F(hkl)|^2$, as amplitudes $|F(hkl)|$ são facilmente encontradas. Porém, nenhuma informação é disponível sobre os ângulos das fases, tornando por isso a descoberta das fases o problema fundamental na determinação da

estrutura de qualquer cristal. Existem quatro métodos principais para determinar as fases do fator de estrutura, sendo:

- Métodos diretos, já usados para pequenas proteínas e em fase de desenvolvimento para proteínas maiores;
- O método de difração anômala com múltiplos comprimentos de onda, que depende da presença de átomos que espalhem forte e irregularmente na própria estrutura da proteína;
- O método de substituição isomórfica, que requer o uso de átomos pesados (átomos com alto número atômico) junto às moléculas de proteína no cristal;
- O método de substituição molecular, que utiliza uma estrutura já resolvida, similar e/ou com enovelamento parecido já determinado.

3.10 Substituição Molecular

Rossmann em 1962, partindo-se do pressuposto que proteínas que tenham estrutura primária semelhantes apresentam o mesmo enovelamento, propôs um método onde uma estrutura conhecida homóloga é ajustada na cela unitária de outra estrutura que não possui informações plausíveis sobre suas fases. Dessa forma, as fases da estrutura conhecida são usadas como uma primeira aproximação para as fases da nova estrutura. Esse método é hoje conhecido como Substituição Molecular.

Nesse método, o ajuste do modelo se faz em uma procura em seis dimensões, ou graus de liberdade, que consistem basicamente em procuras de ordem rotacional e translacional. Se realizarmos as duas buscas concomitantemente terá um alto custo computacional. Entretanto, dividindo-se em busca de rotação primeiro e em seguida em

busca de translação, têm-se três graus de liberdade para o cálculo de rotação e três graus de liberdade para cálculo de translação, diminuindo assim, consideravelmente, o custo computacional.

O modelo inicial (também chamado modelo de busca) a ser usado na procura deve ser o mais homólogo possível da macromolécula a ser estudada. No caso de proteína, uma identidade igual ou superior a 50% deve ser suficiente para a solução dessa etapa. O maior problema com a utilização de modelos com baixa identidade está na estimativa grosseira das fases do modelo a ser estudado, o que pode causar problemas durante a etapa de refinamento da estrutura a ser resolvida (Nolasco, 2005).

Nem sempre é possível aplicar o método da Substituição Molecular, mesmo quando a semelhança entre a estrutura a ser resolvida e o modelo de busca é alta (acima de 50%). As principais causas são:

- Diferenças devido ao empacotamento cristalino,
- Diferenças devido ao estado oligomérico,
- Baixa qualidade estrutural do modelo de busca,
- Baixa qualidade dos dados de difração de raios X.

4. Resultados e Discussão

4.1 *Canavalia marítima* lectin (CML) e *Canavalia gladiata* lectin (CGL)

Tanto a lectina de *Canavalia marítima* quanto a lectina de *Canavalia gladiata* são estruturalmente análogas, com apenas três resíduos diferentes em suas estruturas primárias: na estrutura da CGL aparecem os resíduos Asn²¹, Met¹²⁹ e Pro²⁰², enquanto que na estrutura da CML aparecem os resíduos Ser²¹, Val¹²⁹ e Ser²⁰². A partir do exposto, apresentaremos os resultados conjuntamente das duas proteínas, pois o processo de cristalização e resolução de estrutura de ambas ocorreu de forma semelhante.

As proteínas *Canavalia marítima* e *Canavalia gladiata* (ambas em concentração molar de 5 mg/ml) foram dissolvidas, cada uma, em uma solução tampão (Tris Hcl, pH 7,0) contendo íons⁴ de Cálcio (Ca²⁺) e Manganês (Mn²⁺). Em cada preparação foi adicionado o açúcar man 1-6 man (manose 1-6 manose), com concentração molar de 2 mg/ml, e a mistura foi centrifugada durante 10 minutos a 10.000 rpm.

Inicialmente, procura-se encontrar uma condição de cristalização possível para a obtenção de cristais da proteína. Nesse processo foi utilizado um kit de cristalização denominado **Screen Index** (Hampton Research), que contém 96 soluções previstas pelo método da matriz esparsa. Ao ser encontrado micro-cristais em uma das soluções (solução n° 64 contendo: Sais: Cloreto de níquel, cloreto de cobalto, cloreto de magnésio e cloreto de cádmio, todos com 0,005M; precipitante: polietileno glicol (3350) a 12% e tampão Tris Hcl (pH 8,0)), teve início o processo de otimização, que visa maximizar as condições previstas pela solução de cristalização, para então obter

⁴ Uma vez ligada (reversivelmente) a metais, o sítio ativo de lectinas de leguminosas assumi conformação mais favorável para complexar o ligante.

cristais com tamanho e volume apropriados para a coleta de dados de difração de raios X. Para a otimização da condição prevista, foram realizadas variações de precipitante, de sal e de pH, mantendo fixa a concentração de proteína e do volume do poço (1 ml).

Os cristais foram cristalizados segundo condições contendo:

Sais	Cloreto de Níquel	0,005M
	Cloreto de Magnésio	0,005M
	Cloreto de Cobalto	0,005M
	Cloreto de Cádmio	0,005M
Precipitante	Polietileno Glicol (3350)	9%
Tampão	Tris Hcl	pH 7,5

Esses cristais atingem tamanhos e volumes apropriados para a coleta de dados de difração de raios X, em aproximadamente 48 horas. Os cristais apresentaram dimensões de aproximadamente $0,5 \times 0,5 \times 0,4$ mm para a CML e $0,4 \times 0,6 \times 0,5$ mm para a CGL. Geralmente, esses cristais são frágeis e sensíveis e qualquer movimento brusco realizado na placa pode acarretar a dissolução dos mesmos. Abaixo segue a foto de duas gotas contendo cristais da CML e da CGL, respectivamente (Figura 18).

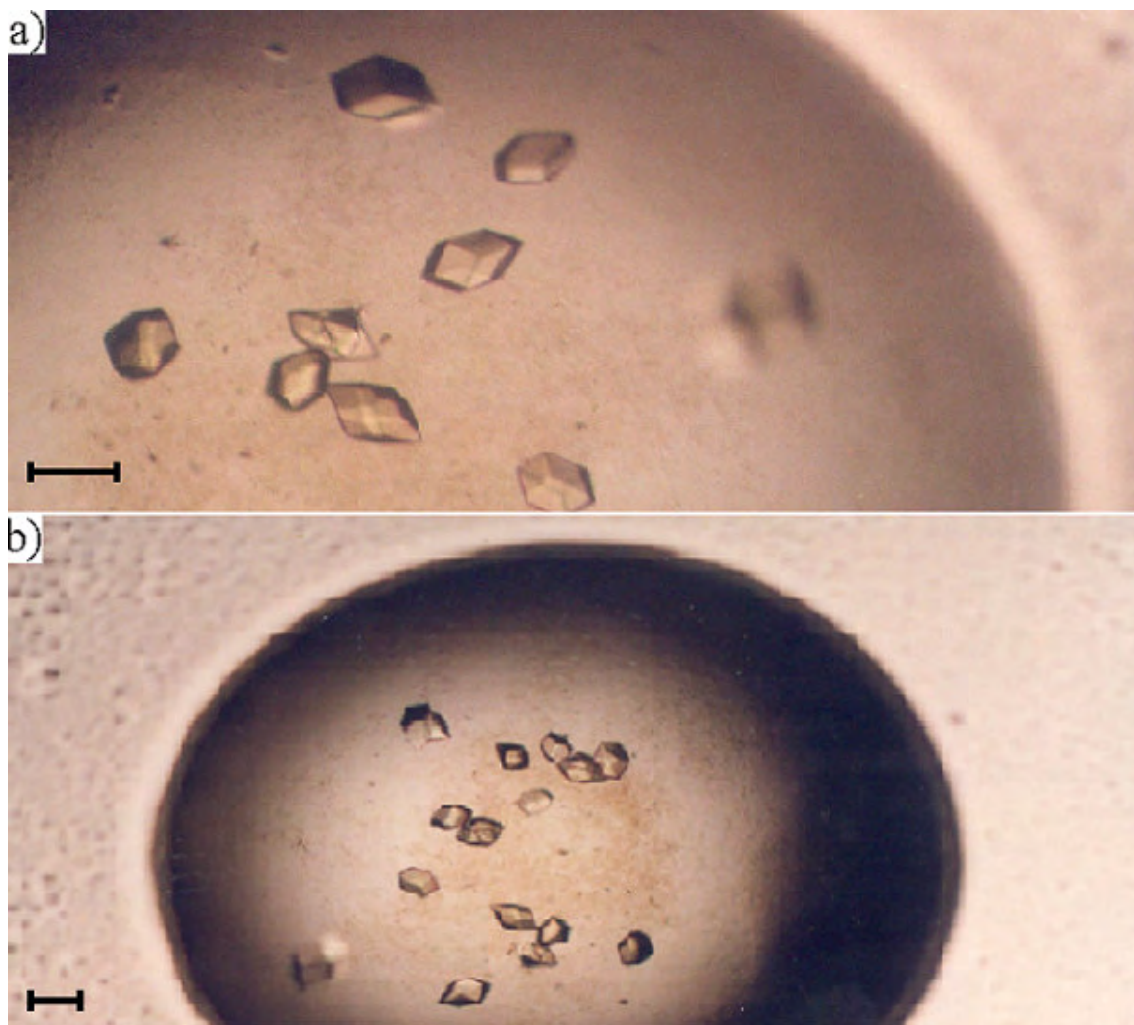


Figura 18: Gotas contendo cristais da proteína CML em a) e cristais da proteína CGL em b).

Dados de difração de raios X (comprimento de onda $\lambda = 1,4270 \text{ \AA}$) de cristais das duas proteínas foram coletados, separadamente, utilizando a fonte de radiação de luz síncrotron, na estação CPr do Laboratório Nacional de Luz Síncrotron, LNLS, localizado na cidade de Campinas/SP, Brasil, utilizando um detector CCD (MARCCD) a uma distância de 50 mm do cristal. Cada cristal foi montado em **loops** e exposto a um fluxo contínuo de nitrogênio a 104K, utilizando uma solução crioprotetora com glicerol 50%. Foram coletadas 160 imagens de cada cristal, com um ângulo de oscilação ϕ igual a 1° e tempo de exposição ao feixe de radiação de 1 minuto por imagem. Os dados

foram processados a partir do programa DENZO (Gewirth, 1995) e escalonados utilizando o programa SCALAPACK (Gewirth, 1995).

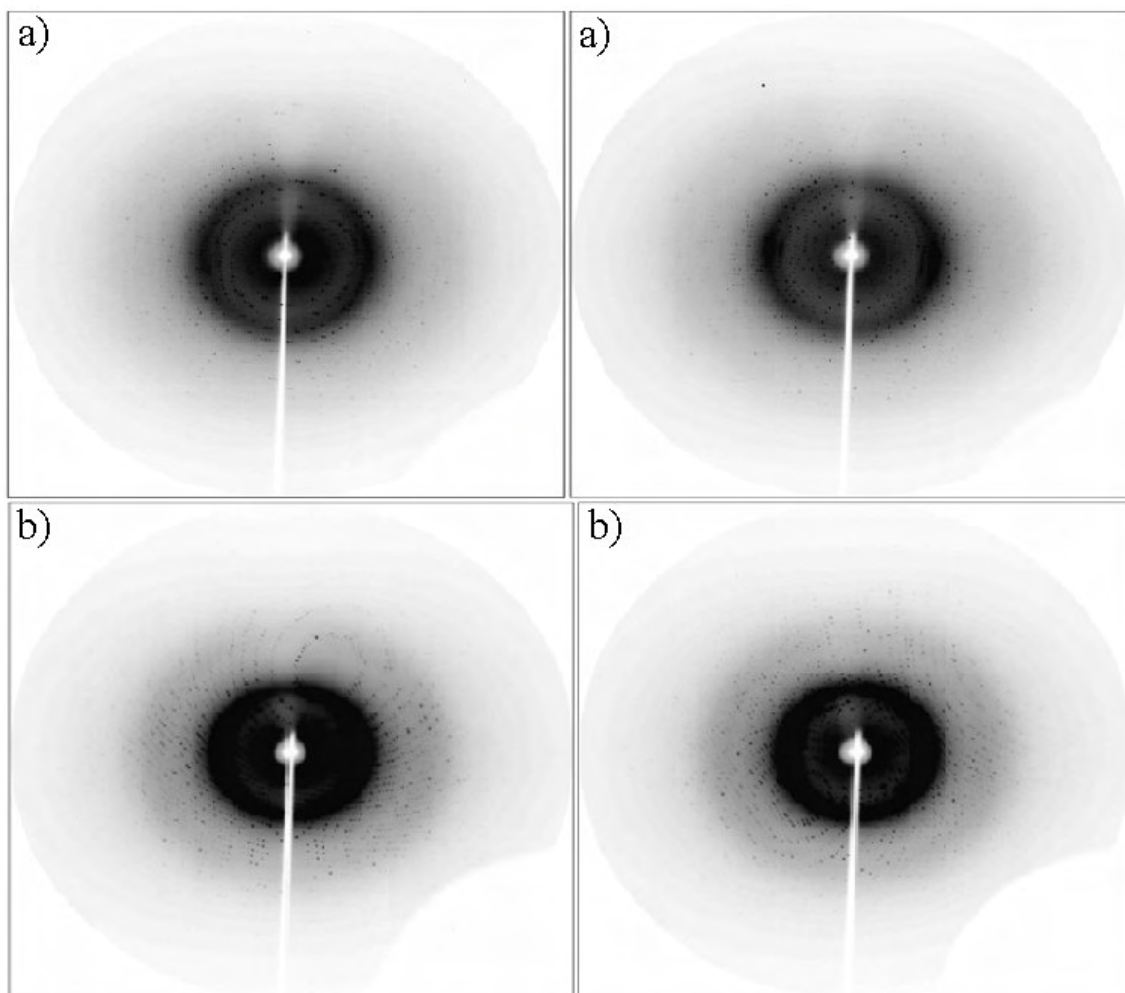


Figura 19: Padrões de difração da proteína CML complexada com o açúcar man 1-6 man em a) e padrões de difração da proteína CGL complexada com o açúcar man 1-6 man em b).

Os dados de difração de raios X dos cristais da proteína CML foram coletados a 1,65 Å de resolução, e da proteína CGL a 1,58 Å. No entanto, quando todos os dados são indexados mostraram pertencer ao grupo de Laue primitivo ortorrômbico $P2_12_12$

(ângulos $\alpha = \beta = \gamma = 90^\circ$), com parâmetros de cela **a** = 59,115 Å, **b** = 80,826 Å e **c** = 99,655 Å para a CML, e parâmetros de cela **a** = 59,233 Å, **b** = 81,243 Å e **c** = 100,064 Å para a CGL. Os valores finais foram obtidos após o refinamento destes parâmetros e utilizando as 160 imagens coletadas de cada proteína.

Tabela 02: Estatística do processamento de dados de difração de raios X da lectina CML e da lectina CGL

Complexos	Lectina (CML)	Lectina (CGL)
a (Å)	59,115	59,233
b (Å)	80,826	81,243
c (Å)	99,655	100,064
α (°)	90	90
β (°)	90	90
γ (°)	90	90
Grupo espacial	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Números de imagens utilizadas	160	160
Mosaicidade	0,25	0,27

Analisando as 160 imagens coletadas de cada proteína, no processo de difração de raios X, podemos observar 53.566 reflexões independentes, completeza de 97,1 (%) e R_{merge} de 4 (%) para a proteína CML, e completeza de 99,8 (%) e R_{merge} de 4,5 (%) para a proteína CGL, sendo 63.347 reflexões independentes. Segue abaixo as tabelas 3.a) e 3.b) com estatísticas detalhadas por faixa de resolução.

Tabela 03.a: Estatística detalhada por faixa de resolução da lectina CML.

Faixa de resolução (Å)	Números de reflexões independentes	Completeza (%)	R _{merge} (%)
99,50 – 4,38	3347	85,9	2,1
4,38 – 3,43	3510	95,8	2,2
3,43 – 2,97	3595	98,5	2,9
2,97 – 2,68	3603	98,9	3,6
2,68 – 2,47	3586	98,1	4,8
2,47 – 2,31	3591	98,0	6,3
2,31 – 2,19	3556	98,7	8,5
2,19 – 2,08	3539	99,3	10,2
2,08 – 2,00	3541	99,5	12,0
2,00 – 1,92	3579	99,1	14,9
1,92 – 1,85	3585	98,5	17,7
1,85 – 1,79	3564	98,6	25,6
1,79 – 1,74	3501	97,9	37,2
1,74 – 1,69	3497	97,6	44,4
1,69 – 1,65	3972	92,8	51,8
Total	53566	97,1	4,0

Completeza = (Números de reflexões observadas / números de reflexões possíveis),

$R_{\text{merge}} = S^{1/2} I - \langle I \rangle^{1/2} / S I$; onde I é a intensidade de uma reflexão e $\langle I \rangle$ é o valor médio da intensidade desta reflexão.

Tabela 03.b: Estatística detalhada por faixa de resolução da lectina CGL.

Faixa de resolução (Å)	Números de reflexões independentes	Completeza (%)	R _{merge} (%)
99,50 – 4,38	3985	99,3	2,2
4,38 – 3,43	4148	99,4	2,3
3,43 – 2,97	4253	99,5	2,8
2,97 – 2,68	4221	99,7	3,5
2,68 – 2,47	4213	99,8	5,1
2,47 – 2,31	4225	99,9	6,9
2,31 – 2,19	4228	99,9	7,9
2,19 – 2,08	4211	99,9	9,3
2,08 – 2,00	4217	99,9	12,2
2,00 – 1,92	4231	99,9	14,1
1,92 – 1,85	4233	99,9	19,5
1,85 – 1,79	4238	99,9	24,2
1,79 – 1,74	4226	99,9	39,3
1,74 – 1,69	4222	99,9	47,7
1,69 – 1,65	4496	99,8	50,3
Total	63347	99,8	4,5

Completeza = (Números de reflexões observadas / números de reflexões possíveis),

$R_{\text{merge}} = S^{1/2} I - \langle I \rangle^{1/2} / S I$; onde I é a intensidade de uma reflexão e $\langle I \rangle$ é o valor médio da intensidade desta reflexão.

4.2 Substituição Molecular

Partindo da observação de que moléculas com seqüências primárias homólogas possuem enovelamentos similares, o conhecimento estrutural dessas moléculas permite simular um modelo inicial do cristal a ser resolvido. A partir deste modelo, pode-se obter as informações necessárias sobre a fase da radiação difratada pelo cristal real (Rhodes, 1999).

Para substituição molecular foi usado o programa AMoRe (Navaza, 1994), integrante do pacote CCP4 (Collaborative Computational Project n° 4; CCP4, 1994). A estrutura utilizada como modelo de busca para substituição molecular, tanto da proteína CML quanto da proteína CGL, foi a estrutura Nativa da proteína *Canavalia gladiata* (código pdb 1wuv). A figura 20 mostra o alinhamento da sequência primária entre as duas lectinas:

Alinhamento das estruturas primárias da CML e da CGL.

Gladiata	ADTIVAVELDTYPNTDIGDPNYPHIGIDIKSVRSKKTAKWNMQNGKVGTAAHIIYNSVGKR
Marítima	ADTIVAVELDTYPNTDIGDPNYPHIGIDIKSVRSKKTAKWNMQNGKVGTAAHIIYNSVGKR
	*****.*****
Gladiata	LSAVVSYPNGDSATVSYDVLDNVLPEWVRVGLSASTGLYKETNTILSWSFTSKLKSNST
Marítima	LSAVVSYPNGDSATVSYDVLDNVLPEWVRVGLSASTGLYKETNTILSWSFTSKLKSNST

Gladiata	HETNALHFMFNQFSKDQKDLILQGDAITGTDGNLELTRVSSNGSPQGSSVGRALFYAPVH
Marítima	HETNALHFVFNQFSKDQKDLILQGDAITGTDGNLELTRVSSNGSPQGSSVGRALFYAPVH
	*****:
Gladiata	IWESSAVVASFDATFTFLIKSPDSHPADGIAFFISNIDSSIPSGSTGRLLGLFPDAN
Marítima	IWESSAVVASFDATFTFLIKSDSHPADGIAFFISNIDSSIPSGSTGRLLGLFPDAN
	*****.

Figura 20: Alinhamento das estruturas primárias da CML e da CGL com 99% de identidade.

Os dados de difração de raios X escalonados, utilizando o programa SCALEPACK (arquivo.sca), tiveram que ser convertidos para o formato específico do pacote CCP4 (arquivo.mtz), utilizando o programa F2MTZ (integrante do pacote CCP4).

As funções de rotação foram calculadas para três diferentes faixas de resolução, utilizando a rotina **Roting**, nos intervalos de resolução 8,5 – 4,0Å, 7,0 – 2,5Å e 5,5 – 2,5Å, com passos de 2° e raio de integração de 20Å. Os melhores resultados (maior valor para o coeficiente de correlação e menor R_{factor}) foram utilizados para o cálculo da translação, através da rotina **Traing**. O melhor resultado para a proteína CML foi encontrado na faixa de resolução 5,5 – 2,5Å, com coeficiente de correlação de 55,7% e R_{factor} de 43,4%, conforme apresenta a Tabela 04.a. Para a proteína CGL, o melhor resultado foi encontrado na mesma faixa de resolução, porém com coeficiente de correlação de 58,3% e R_{factor} de 41,2%, conforme a Tabela 04.b.

Tabela 04.a: Ângulos de rotação, distâncias de translação, coeficiente de correlação (CC) e R_{factor} obtidos na substituição molecular do cristal da lectina CML

Resolução (Å)	a (°)	b (°)	c (°)	T_x	T_y	T_z	CC (%)	R_{factor} (%)
8,5 – 4,0Å	81,02	75,36	153,49	0,0142	0,0098	0,0105	53,5	44,1
7,0 – 2,5Å	81,08	75,97	152,30	0,0163	0,0103	0,0101	55,6	43,8
5,5 – 2,5Å	81,13	76,01	152,17	0,0169	0,0123	0,0100	55,7	43,4

Tabela 04.b: Ângulos de rotação, distâncias de translação, coeficiente de correlação (CC) e R_{factor} obtidos na substituição molecular do cristal da lectina CGL

Resolução (Å)	a (°)	b (°)	c (°)	T_x	T_y	T_z	CC (%)	R_{factor} (%)
8,5 – 4,0Å	82,61	74,29	133,22	0,0153	0,0113	0,0097	55,9	42,8
7,0 – 2,5Å	82,68	74,33	134,01	0,0167	0,0124	0,0111	58,0	41,7
5,5 – 2,5Å	82,75	74,91	133,87	0,0172	0,0129	0,0105	58,3	41,2

As funções de rotação e translação foram novamente calculadas para quatro grupos espaciais diferentes, pertencentes ao sistema primitivo ortorrômico, na faixa de resolução de 5,5 – 2,5Å, a fim de descobrir qual desses grupos apresentam o melhor resultado. Após a análise foi confirmada a hipótese inicial de que o grupo espacial correto é o P2₁2₁2 (Tabela 05).

Tabela 05.a: Ângulos de rotação, distâncias de translação, coeficiente de correlação (CC) e R_{factor} obtidos para quatro grupos espaciais do sistema ortorrômico (CML)

Grupo Espacial	a (°)	B (°)	c (°)	T _x	T _y	T _z	CC (%)	R _{factor} (%)
P222	81,12	76,01	153,17	0,8399	0,3561	0,1002	34,2	51,0
P222 ₁	81,12	76,00	152,16	0,3222	0,0139	0,2564	38,3	48,3
P2 ₁ 2 ₁ 2	81,13	76,01	152,17	0,0169	0,0123	0,0100	55,7	43,4
P2 ₁ 2 ₁ 2 ₁	81,17	76,02	152,23	0,0945	0,0124	0,4350	40,9	47,9

Tabela 05.b: Ângulos de rotação, distâncias de translação, coeficiente de correlação (CC) e R_{factor} obtidos para quatro grupos espaciais do sistema ortorrômico (CGL)

Grupo Espacial	a (°)	b (°)	c (°)	T _x	T _y	T _z	CC (%)	R _{factor} (%)
P222	82,73	74,91	134,27	0,5049	0,1971	0,0902	36,7	52,8
P222 ₁	82,74	74,92	133,86	0,1963	0,0134	0,3703	40,1	47,1
P2 ₁ 2 ₁ 2	82,75	74,91	133,87	0,0172	0,0129	0,0105	58,3	41,2
P2 ₁ 2 ₁ 2 ₁	82,78	74,93	133,89	0,0823	0,0117	0,5199	47,5	45,3

4.3 Refinamento Cristalográfico

O refinamento cristalográfico das estruturas protéicas foi realizado utilizando o programa REFMAC5 (MURSHUDOV et al., 1997), parte integrante do pacote CCP4. Inicialmente, para cada proteína, foi realizado o refinamento de corpo rígido utilizando 20 ciclos, e em seguida foi calculado o mapa de densidade eletrônica ($2F_{\text{obs}} - F_{\text{calc}}$) da estrutura. Utilizando o programa XtalView (McREE, 1999), foi feita a reconstrução manual da proteína CML, substituindo os três resíduos diferentes (Asn²¹, Met¹²⁹ e Pro²⁰²), presentes no modelo de busca, pelos resíduos correspondentes (Ser²¹, Val¹²⁹ e Ser²⁰²), conforme observado no alinhamento das estruturas.

Após a substituição dos resíduos foi feito mais um refinamento de corpo rígido e reconstrução manual, checando cada resíduo e os posicionando dentro da sua densidade eletrônica. Posteriormente, foi feito outro refinamento (**Restrained refinement**), agora do complexo proteína/ligante (man 1-6 man). Novamente foi checado o modelo corrigindo a posição de resíduos que estejam fora da densidade eletrônica, posicionando-se, em seguida, o açúcar corretamente dentro de sua própria densidade eletrônica.

Na etapa seguinte, foi realizado um novo passo de refinamento (**Unrestrained refinement**), onde refinou-se os fatores de temperatura isotrópicos e anisotrópicos. Em seguida, foram incluídas 396 moléculas de água no modelo para a CML e 398 moléculas de água para a CGL e foi feita nova checagem manual, para verificar se as moléculas de água estão realmente encaixadas dentro de sua densidade eletrônica. Um novo refinamento do fator de temperatura isotrópico e anisotrópico foi realizado.

Durante a etapa de refinamento cristalográfico, além das 396 moléculas de água adicionadas ao modelo dois sítios de ligação de açúcar foram preenchidos. No final

desse processo, um novo passo de refinamento considerando simetria não cristalográfica foi realizado. Assim, a estrutura contendo 474 aminoácidos foi refinada e ao final desta etapa apresentou R_{factor} de 18,67% e R_{free} de 22,35% para a CML e R_{factor} de 19,82% e R_{free} de 22,67% para a CGL.

Os valores obtidos na etapa de refinamento cristalográfico são mostrados na tabela (Tabela 06):

Tabela 06: Estatística do refinamento cristalográfico da proteína CML e da Proteína CGL

Proteínas	CML	CGL
Faixa de resolução utilizada (Å)	99,50 – 1,65	100 – 1,58
Nº de reflexões utilizadas no refinamento	53566	63347
Números de átomos não hidrogênio	4029	4049
Números de resíduos (dímero)	474	474
Números de moléculas de água (dímero)	396	398
Nº de átomos de Cálcio ligados (dímero)	2	2
Nº de átomos de Manganês ligados (dímero)	3	3
R_{sym} (%) *	5,7	5,3
R_{factor} final (%) **	18,67	19,82
R_{free} final (%) ***	22,35	22,67
Coeficiente de correlação final $F_{\text{obs}} - F_{\text{calc}}$ (%)	95,8	95,5
Coeficiente de correlação livre (%)	94,0	94,0
Grupo espacial	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Multiplicidade	2,6	2,5

* $R_{\text{sym}} = 100 \sum |I(h) - \langle I(h) \rangle| / \sum I(h)$ com $I(h)$, intensidade observada e $\langle I(h) \rangle$, intensidade de reflexões h sobre todas as medidas de $I(h)$.

** $R_{\text{factor}} = 100 \times \sum (|F_{\text{obs}} - F_{\text{calc}}|) / \sum (F_{\text{obs}})$, soma obtidas sobre todas as reflexões.

*** $R_{\text{free}} = R_{\text{factor}}$ para 10% dos dados, os quais não foram incluídos no refinamento.

4.4 Descrições da Estrutura

As lectinas CML e CGL, quando complexadas com o ligante manose 1-6 manose, foram refinadas considerando uma unidade assimétrica dimérica. Entretanto não se pode afirmar que essa seja a estrutura da proteína em solução na sua forma nativa, pois, nas outras estruturas já determinadas destas lectinas em complexo com outros açúcares (PDB: 2ow4, 2p37, 2p34, 2ovu, 2ef6 e 2p2k), não foram realizadas estudos que indicassem sua forma nativa em solução. A única exceção é a estrutura da lectina CGL complexada com o ligante manose 1-4 manose onde foi determinada a forma tetramérica para o complexo em solução.

As estruturas calculadas são compostas por 237 resíduos de aminoácidos por monômero. No caso da CML os monômeros são compostos exclusivamente por nove fitas beta (**front face**) e oito fitas beta (**back face**), para o monômero A, e nove fitas beta (**front face**) e sete fitas beta (**back face**), para o monômero B (Figura 21.a). Já para a lectina CGL os monômeros são compostos exclusivamente por nove fitas beta (**front face**) e seis fitas beta (**back face**), tanto para o monômero A quanto para o B (Figura 21.b). Todos os monômeros de ambas as proteínas são interligadas por voltas e alças formando uma estrutura na forma achatada.

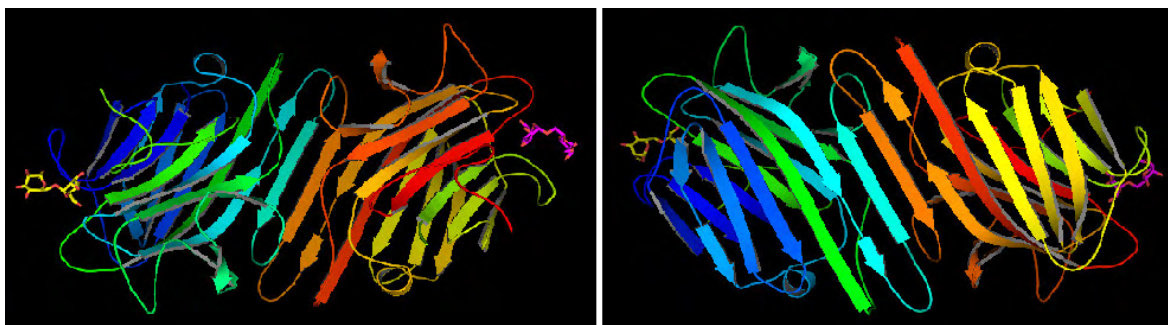


Figura 21.a: Diagrama do dímero da lectina CML na perspectiva **front face** e **back face** respectivamente. Programa utilizado Pymol (DELANO, 2004). Representação Ribbon Cartoon.

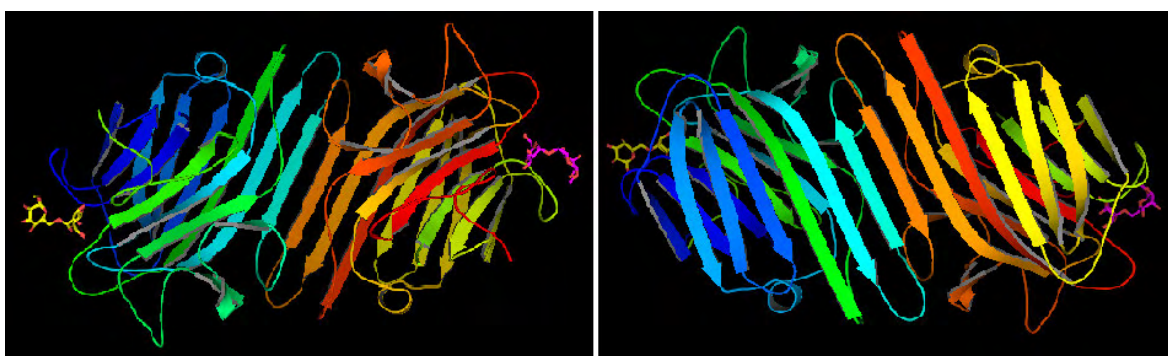


Figura 21.b: Diagrama do dímero da lectina CGL na perspectiva **front face** e **back face** respectivamente. Programa utilizado Pymol (DELANO, 2004). Representação Ribbon Cartoon.

Esse tipo de enovelamento é conhecido como “ β -sanduíche”, por ser composto de folhas betas antiparalelas (cerca de 60%). A orientação antiparalela das folhas beta, e a mudança entre uma face e outra, forma uma estrutura rígida muito forte. Foi encontrado apenas uma α -hélice, contendo uma única volta, formada pelos resíduos (Leu⁸¹, Asp⁸², Asn⁸³ e Val⁸⁴) em cada monômero.

Quatro **loops** localizados na parte superior dos monômeros formam o sítio de ligação, onde está complexado o açúcar (Figura 22). O sítio de ligação do monômero A da lectina CML é constituído por seis resíduos (Tyr¹², Asn¹⁴, Gly⁹⁸, Leu⁹⁹, Tyr¹⁰⁰ e Arg²²⁸), que formam pontes de hidrogênio com o açúcar⁵. Outros dois resíduos (Asp¹⁶ e

⁵ Dois desses resíduos (Asn¹⁴ e Arg²²⁸) formam possíveis pontes de hidrogênio com uma molécula de água através dos átomos ND2 e NH1, respectivamente.

Thr²²⁶) formam pontes de hidrogênio com duas moléculas de água, que estão estabelecendo possíveis pontes de hidrogênio diretamente com o açúcar. Apenas uma molécula de água forma ponte de hidrogênio exclusivamente com o açúcar (átomo O3). No total são três moléculas de água que formam pontes de hidrogênio com o ligante (Figura 23).

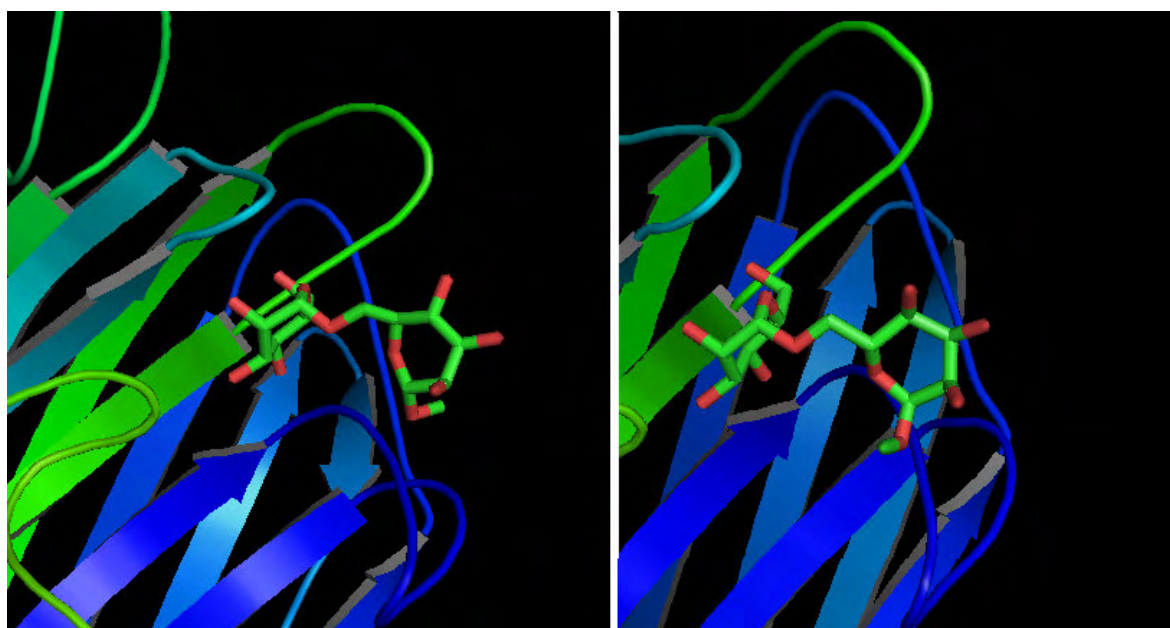


Figura 22: Quatros **loops** localizados na parte superior dos monômeros da lectinas CML e CML respectivamente.

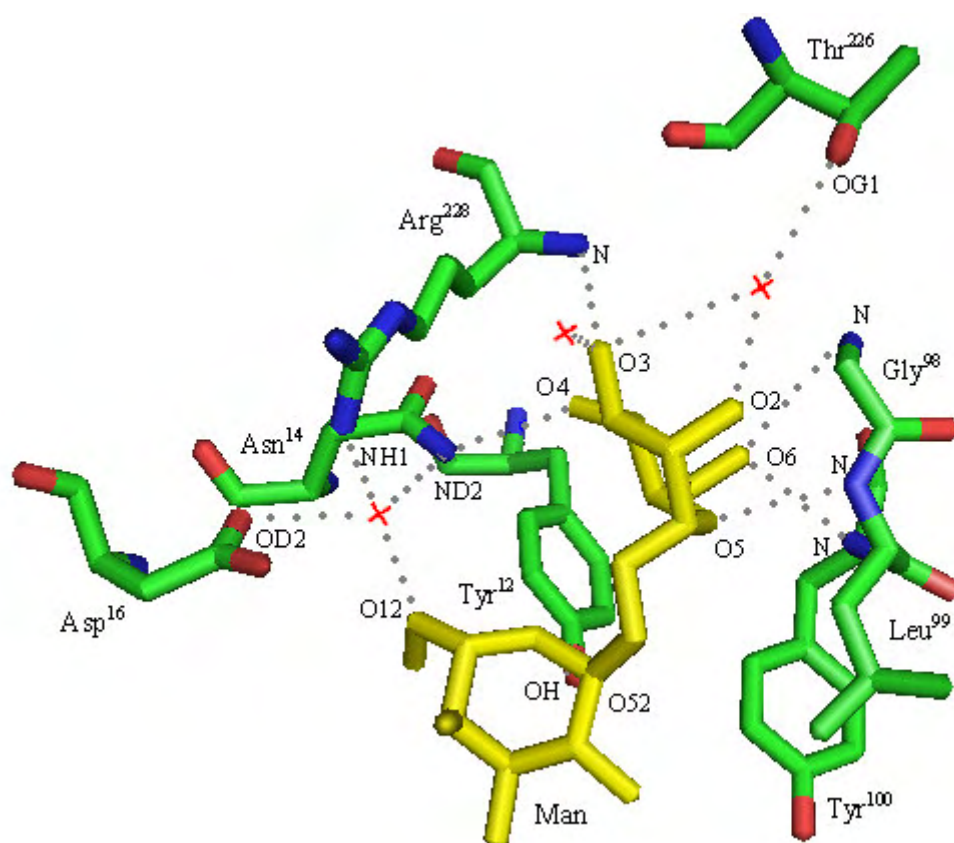


Figura 23: Sítio de ligação do açúcar no monômero A da lectina CML.

Abaixo segue tabela 07 que estabelece todas as medidas das distâncias das possíveis pontes de hidrogênio presentes no sítio de ligação do monômero A da lectina CML.

Tabela 07: Distâncias das possíveis pontes de hidrogênio presentes no sítio de ligação do Monômero A da lectina CML man 1-6 man. Calculado pelo programa XtalView (MCREE, 1999)

	Átomo		Átomo	Distância (Å)
Manose	O2	HOH 292	H	2,53
Manose	O3	HOH 264	H	2,63
Manose	O3	Arg ²²⁸	N	2,84
Manose	O4	Asn ¹⁴	ND2	2,81
Manose	O5	Leu ⁹⁹	N	3,03
Manose	O6	Gly ⁹⁸	N	3,23
Manose	O6	Tyr ¹⁰⁰	N	2,96
Manose	O12	HOH	48	2,70
Manose	O52	Tyr ¹²	OH	3,11
HOH 48	H	Arg ²²⁸	NH1	2,85
HOH 48	H	Asp ¹⁶	OD2	2,71
HOH 48	H	Thr ²²⁶	OG1	2,84

O sítio de ligação do monômero B da lectina CML é constituído por seis resíduos (Tyr¹², Asn¹⁴, Gly⁹⁸, Leu⁹⁹, Tyr¹⁰⁰ e Arg²²⁸), formando pontes de hidrogênio diretamente com o açúcar, sendo que dois desses resíduos (Asn¹⁴ e arg²²⁸) formam possíveis pontes de hidrogênio com uma molécula de água. Dois resíduos (Asp¹⁶ e Trh²²⁶) formam pontes de hidrogênio com duas moléculas de água, que formam pontes de hidrogênio diretamente com o açúcar. Apenas duas moléculas de água formam pontes de hidrogênio exclusivamente com o açúcar. No total são quatro moléculas de água que formam pontes de hidrogênio com o ligante (Figura 24).

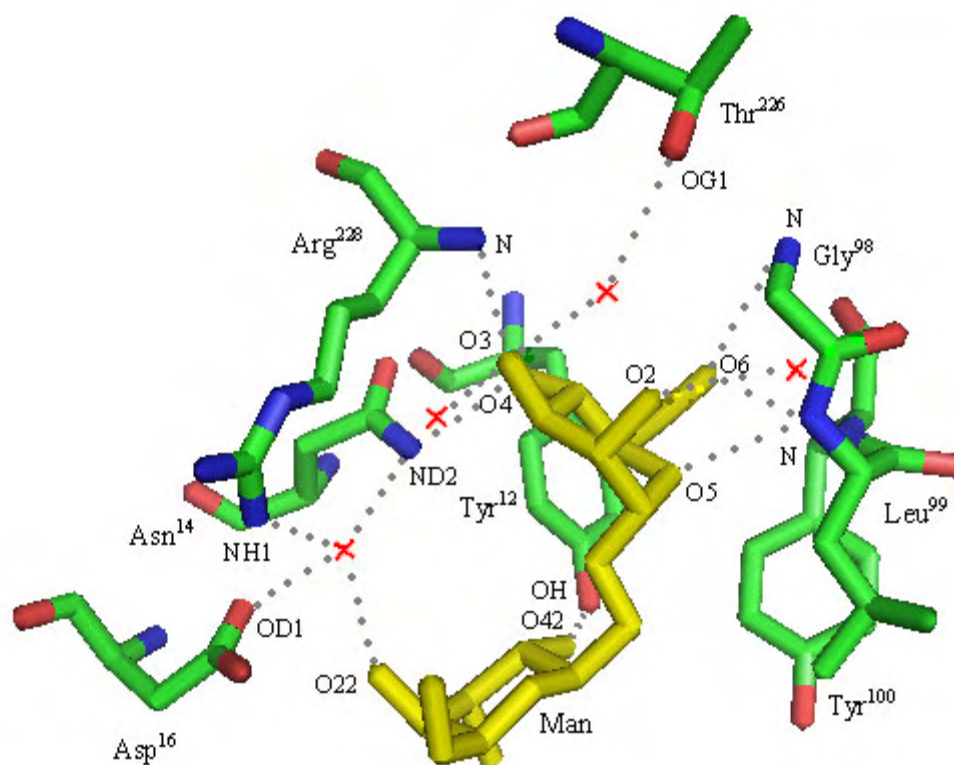


Figura 24: Sítio de ligação do açúcar no monômero B da lectina CML.

Abaixo segue tabela 08 que estabelece todas as medidas das distâncias das possíveis pontes de hidrogênio presentes no sítio de ligação do monômero B da lectina CML.

Tabela 08: Distâncias das possíveis pontes de hidrogênio presentes no sítio de ligação do Monômero B da lectina CML man 1-6 man. Calculado pelo programa XtalView (MCREE, 1999)

	Átomo		Átomo	Distância (Å)
Manose	O2	HOH 352	H	3,21
Manose	O3	HOH 376	H	2,67
Manose	O3	Arg ²²⁸	N	2,78
Manose	O3	HOH 317	H	3,26
Manose	O4	Asn ¹⁴	ND2	2,79
Manose	O5	Leu ⁹⁹	N	3,00
Manose	O6	Gly ⁹⁸	N	3,18
Manose	O6	Tyr ¹⁰⁰	N	2,98
Manose	O22	HOH 88	H	2,61
Manose	O42	Tyr ¹²	OH	2,22
HOH 88	H	Asn ¹⁴	ND2	2,87
HOH 88	H	Asp ¹⁶	OD1	2,54
HOH 88	H	Arg ²²⁸	NH1	2,98
HOH 317	H	Thr ²²⁶	OG1	2,87

Em relação à lectina CGL, o sítio de ligação do monômero A é constituído por seis resíduos (Tyr¹², Asn¹⁴, Gly⁹⁸, Leu⁹⁹, Tyr¹⁰⁰ e Arg²²⁸), que formam pontes de hidrogênio com o açúcar, mais dois resíduos (Asp¹⁶ e Thr²²⁶) formando pontes de hidrogênio com duas moléculas de água, estas que estão estabelecendo possíveis pontes de hidrogênio diretamente com o açúcar. Por último, mais três moléculas de água formam pontes de hidrogênio exclusivamente com o açúcar. No total são seis moléculas de água que formam pontes de hidrogênio com o ligante (Figura 25).

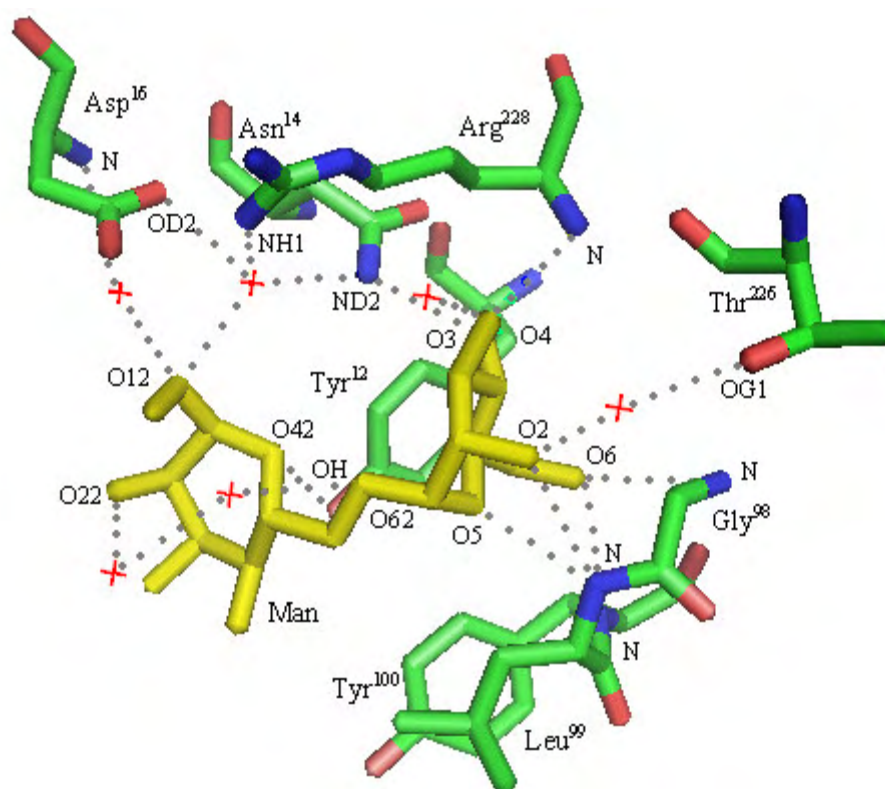


Figura 25: Sítio de ligação do açúcar no monômero A da lectina CGL.

Abaixo segue tabela 09 que estabelece todas as medidas das distâncias das pontes de hidrogênio presentes no sítio de ligação do monômero A da lectina CGL.

Tabela 09: Distâncias das possíveis pontes de hidrogênio presentes no sítio de ligação do Monômero A da lectina CGL man 1-6 man. Calculado pelo programa XtalView (MCREE, 1999)

	Átomo		Átomo	Distância (Å)
Manose	O2	HOH 332	H	2,58
Manose	O3	HOH 142	H	2,72
Manose	O3	Arg ²²⁸	N	2,85
Manose	O4	Asn ¹⁴	ND2	2,83
Manose	O5	Leu ⁹⁹	N	3,03
Manose	O6	Gly ⁹⁸	N	3,24
Manose	O6	Tyr ¹⁰⁰	N	2,98
Manose	O12	HOH 26	H	2,69
Manose	O12	HOH 386	H	3,18
Manose	O22	HOH 346	H	3,20
Manose	O42	Tyr ¹²	OH	2,99
Manose	O62	HOH 163	H	2,96
HOH 26	H	Asn ¹⁴	ND2	2,81
HOH 26	H	Asp ¹⁶	OD2	2,76
HOH 26	H	Arg ²²⁸	NH1	2,83
HOH 332	O	HOH 163	H	2,90
HOH 332	H	Thr ²²⁶	OG1	2,72
HOH 386	H	Asp ¹⁶	N	3,03

O sítio de ligação do monômero B, da mesma lectina CGL, é constituído por seis resíduos (Tyr¹², Asn¹⁴, Gly⁹⁸, Leu⁹⁹, Tyr¹⁰⁰ e Arg²²⁸), formando pontes de hidrogênio diretamente com o açúcar, sendo que os resíduos Asn¹⁴ e Arg²²⁸ formam duas possíveis pontes de hidrogênio (uma diretamente com o açúcar e a outra com uma molécula de água). Apenas o resíduo Asp¹⁶ forma ponte de hidrogênio com uma molécula de água, sendo que esta forma ponte de hidrogênio diretamente com o açúcar. Três moléculas de água formam pontes de hidrogênio exclusivamente com o açúcar. No

total são cinco moléculas de água que formam pontes de hidrogênio com o ligante (Figura 26).

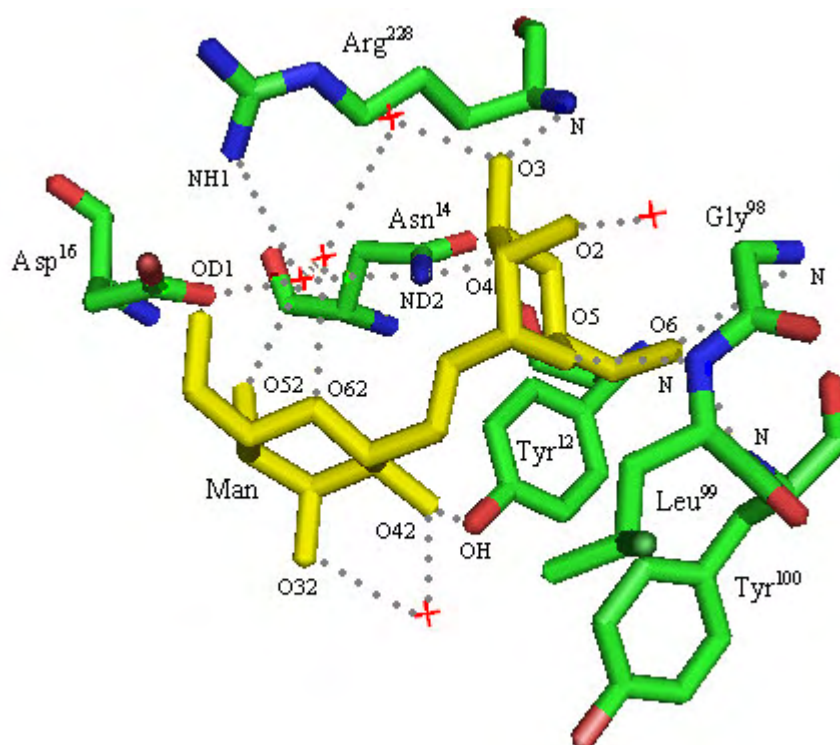


Figura 26: Sítio de ligação do açúcar no monômero B da lectina CGL.

Abaixo segue tabela 10 que estabelece todas as medidas das distâncias das pontes de hidrogênio presentes no sítio de ligação do monômero B da lectina CGL.

Tabela 10: Distâncias das possíveis pontes de hidrogênio presentes no sítio de ligação do Monômero B da lectina CGL man 1-6 man. Calculado pelo programa XtalView (MCREE, 1999)

	Átomo		Átomo	Distância (Å)
Manose	O2	HOH 341	H	2,94
Manose	O3	HOH 257	H	2,64
Manose	O3	Arg ²²⁸	N	2,85
Manose	O4	Asn ¹⁴	ND2	2,77
Manose	O5	Leu ⁹⁹	N	3,13
Manose	O6	Gly ⁹⁸	N	2,96
Manose	O6	Tyr ¹⁰⁰	N	2,98
Manose	O32	HOH 255	H	3,23
Manose	O42	HOH 255	H	2,77
Manose	O42	Tyr ¹²	OH	2,28
Manose	O52	HOH 135	H	2,99
Manose	O62	HOH 135	H	3,23
HOH 45	O	HOH 135	H	3,24
HOH 45	H	Asn ¹⁴	ND2	2,79
HOH 45	H	Asp ¹⁶	OD1	2,59
HOH 45	H	Arg ²²⁸	NH1	3,06
HOH 257	O	HOH	135	3,07

Próximo ao sítio de ligação encontra-se dois átomos, cálcio e manganês, ligados reversivelmente à proteína. Esses metais estão localizados a uma distância de 4.19Å um do outro, para a lectina CML, e a 4,23Å um do outro para a lectina CGL (Figura 27).

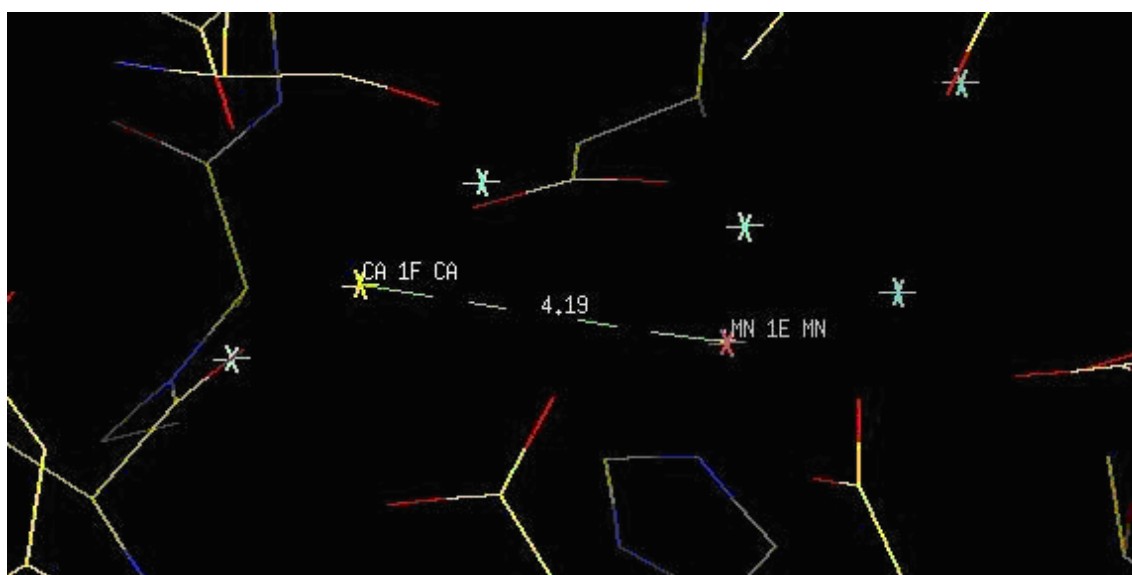


Figura 27: Distância entre o átomo de Cálcio e Manganês presente no interior da proteína CML. Gerado pelo programa XtalView (MCREE, 1999).

Os átomos de cálcio e manganês são coordenados por cadeias laterais constituídas de oito aminoácidos (Glu⁸, Asp¹⁰, Asn¹⁴, Asp¹⁹, His²⁴, Val³², Ser³⁴ e Arg²²⁸), e três moléculas de água (Figura 28). Um terceiro átomo de manganês foi encontrado na superfície do monômero B, de ambas as proteínas, e está coordenado por quatro resíduos (Lys⁵⁹, Asp⁸⁰, Asp⁸² e Asn⁸³) e uma molécula de água. Dois desses resíduos (Asp⁸² e Asn⁸³) estão presentes na única α -hélice das proteínas. Uma das possíveis explicações para o aparecimento desse terceiro átomo de manganês, fora de seu sítio ativo, é o fato das proteínas terem sido supersaturadas com esse tipo de átomo.

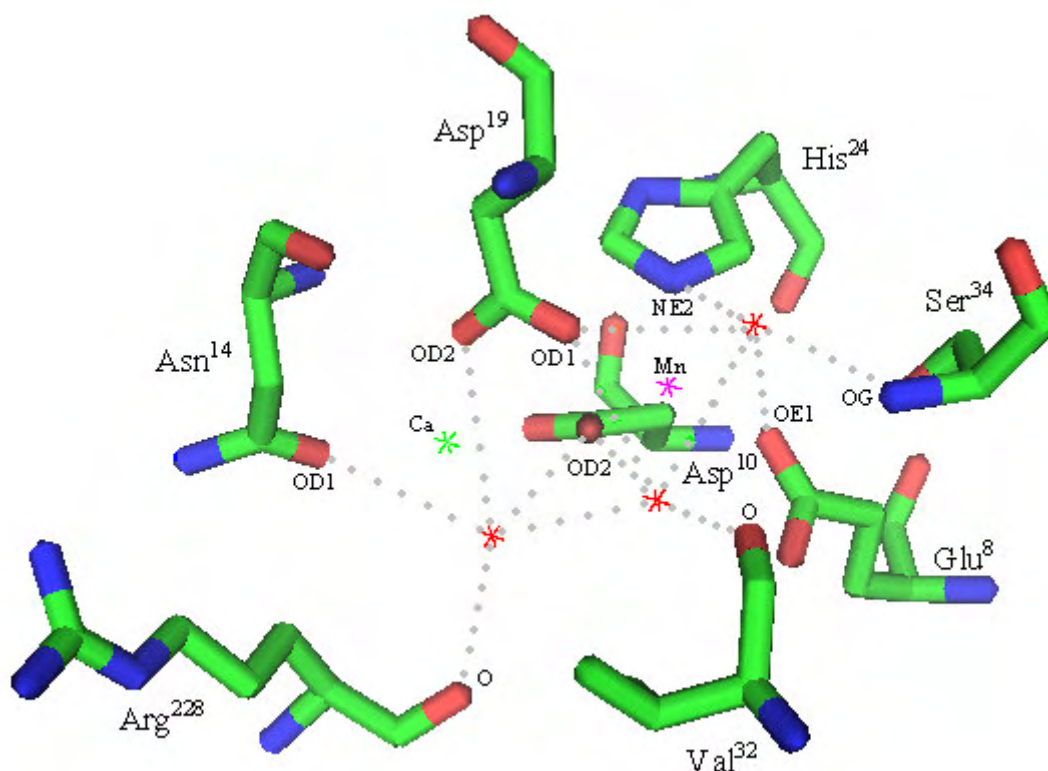


Figura 28: Diagrama do sítio de ligação dos átomos de Cálcio e Manganês da lectina CML. Programa utilizado Pymol (DELANO, 2004).

Em ambos os sítios de ligação, tanto do monômero A quanto do monômero B das duas lectinas, encontra-se presente o resíduo Asp²⁰⁸, que estabelece interação eletrostática com o açúcar presente no sítio de ligação. No monômero A da lectina CML o átomo OD1 da Asp²⁰⁸ está a uma distância de 2,68Å do átomo O6 da manose 1, e o átomo OD2 da Asp²⁰⁸ está a uma distância de 2,55Å do átomo O4 da manose 1. Entretanto, no monômero B, o átomo OD1 da Asp²⁰⁸ está a uma distância de 2,68Å do átomo O6 da manose 1, e o átomo OD2 da Asp²⁰⁸ está a uma distância de 2,60Å do átomo O4 da manose 1.

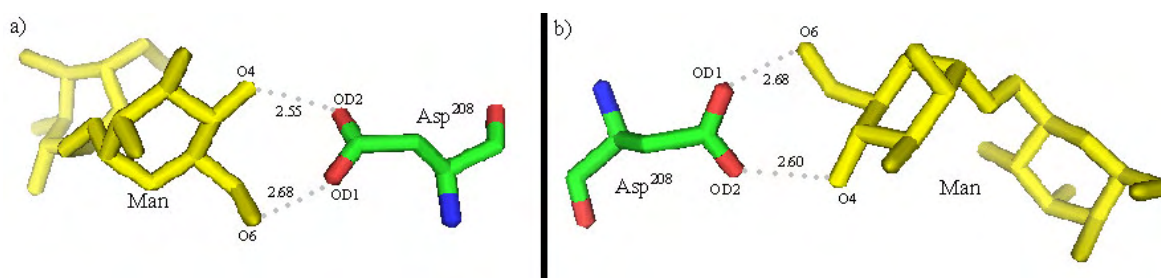


Figura 29: Medida da distância do resíduo Asp²⁰⁸ com a manose 1-6 manose da lectina CML gerada pelo programa XtalView (MCREE, 1999). Figura gerada pelo programa Pymol (DELANO, 2004). a) Monômero A. b) Monômero B.

No monômero A da lectina CGL, o átomo OD1 da Asp²⁰⁸ está a uma distância de 2,67Å do átomo O6 da manose 1, e o átomo OD2 da Asp²⁰⁸ está a uma distância de 2,54Å do átomo O4 da manose 1. Entretanto, no monômero B da mesma lectina o átomo OD1 da Asp²⁰⁸ está a uma distância de 2,75Å do átomo O6 da manose 1, e o átomo OD2 da Asp²⁰⁸ está a uma distância de 2,61Å do átomo O4 da manose 1.

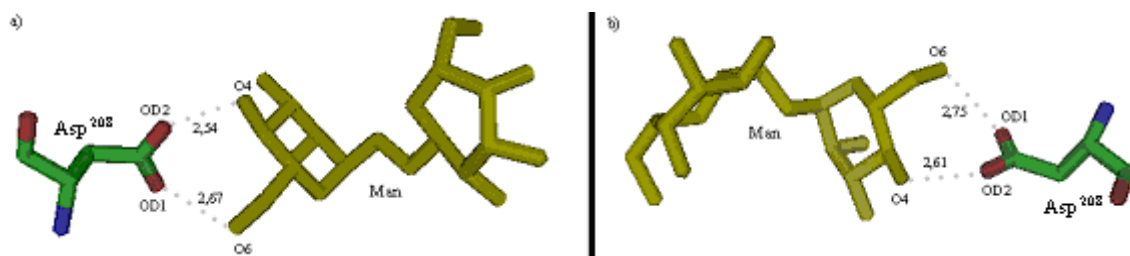


Figura 30: Medida da distância do resíduo Asp²⁰⁸ com a manose 1-6 manose da lectina CGL gerada pelo programa XtalView (MCREE, 1999). Figura gerada pelo programa Pymol (DELANO, 2004). a) Monômero A. b) Monômero B.

4.5 Qualidade do modelo

A estrutura cristalográfica do dímero da lectina CML man 1-6 man apresenta 455 resíduos (96,8%) nas regiões mais favoráveis do diagrama de Ramachandran Φ - Ψ , 13 resíduos (2,8%) nas regiões adicionalmente permitidas e apenas 2 resíduos (0,4%) em regiões não permitidas do diagrama (figura 231.a). Os resíduos Ser¹¹⁹ do monômero A e Thr¹²⁰ do monômero B estão presentes em regiões não permitidas no diagrama de Ramachandran.

Já para estrutura cristalográfica do dímero da lectina CGL man 1-6 man, foi apresentado 456 resíduos (97%) nas regiões mais favoráveis do diagrama de Ramachandran Φ - Ψ , 11 resíduos (2,3%) nas regiões adicionalmente permitidas e apenas 3 resíduos (0,6%) em regiões não permitidas do diagrama (figura 31.b). Os resíduos Asn¹¹⁸ e Glu¹²² do monômero A e Thr¹²⁰ do monômero B estão presentes em regiões não permitidas no diagrama de Ramachandran. As análises do mapa de densidade eletrônica ($2F_{\text{obs}} - F_{\text{calc}}$) levam a crer na veracidade da estrutura, visto que todos os resíduos de aminoácidos se ajustam perfeitamente com suas respectivas densidades eletrônicas. Esses resíduos estão presentes na superfície da proteína.

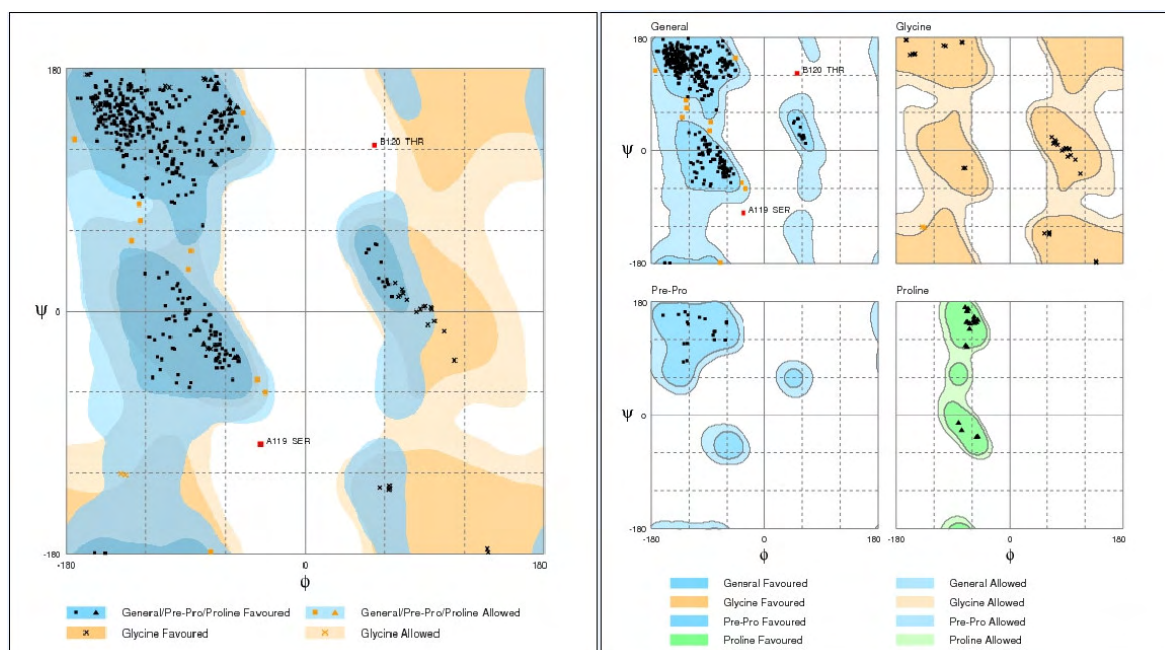


Figura 31.a: Diagrama de Ramachandran Φ - Ψ para a estrutura CML man 1-6 man. Imagem gerada utilizando o webserve RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>).

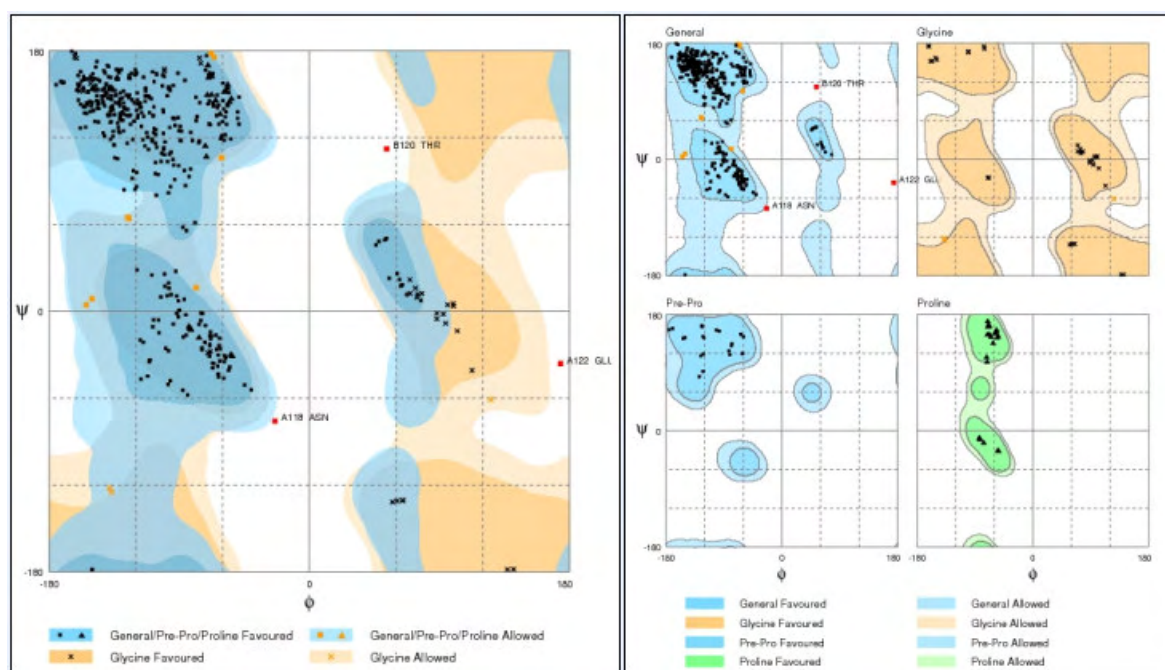


Figura 31.b: Diagrama de Ramachandran Φ - Ψ para a estrutura CGL man 1-6 man. . Imagem gerada utilizando o webserve RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>).

Posteriormente foi feita a comparação das estruturas, através da sobreposição da estrutura da lectina CML nativa (estrutura depositada no PDB – Protein Data Bank, código de acesso 2cwm) com a estrutura da lectina CML complexada com o açúcar man 1-6 man. Para a sobreposição das estruturas foi utilizado o programa LSQKAB, componente do pacote CCP4 (Collaborative Computational Project nº 4, 1994). Após a sobreposição foi calculado o RMSD da estrutura obtendo, assim, a quantização desta comparação. O mesmo processo foi feito para a estrutura da lectina CGL nativa (estrutura depositada no PDB – Protein Data Bank, código de acesso 1wuv) com a estrutura da lectina CGL complexada com o açúcar man 1-6 man. Apesar dos diferentes grupos espaciais, C222₁ para a proteína CGL nativa e P2₁2₁2 para a proteína CGL complexada ao ligante man 1-6 man, a sobreposição das estruturas revelou grande semelhança entre elas.

Para o cálculo do RMSD individual de cada resíduo, de ambas as proteínas, foi utilizado o programa Profit. Através desse programa foi feito o cálculo do RMSD de todos os átomos presente na estrutura, ou seja, foi calculado o RMSD médio de 1,218 Å para o dímero da CML e RMSD médio de 0,834 Å para o dímero da CGL. Dessa forma, foi construída uma planilha para cada lectina, onde consta o resultado do cálculo para todos os resíduos. Posteriormente esses valores foram plotados no programa Origin para poder identificar os picos e saber quais as regiões de maior variação de RMSD.

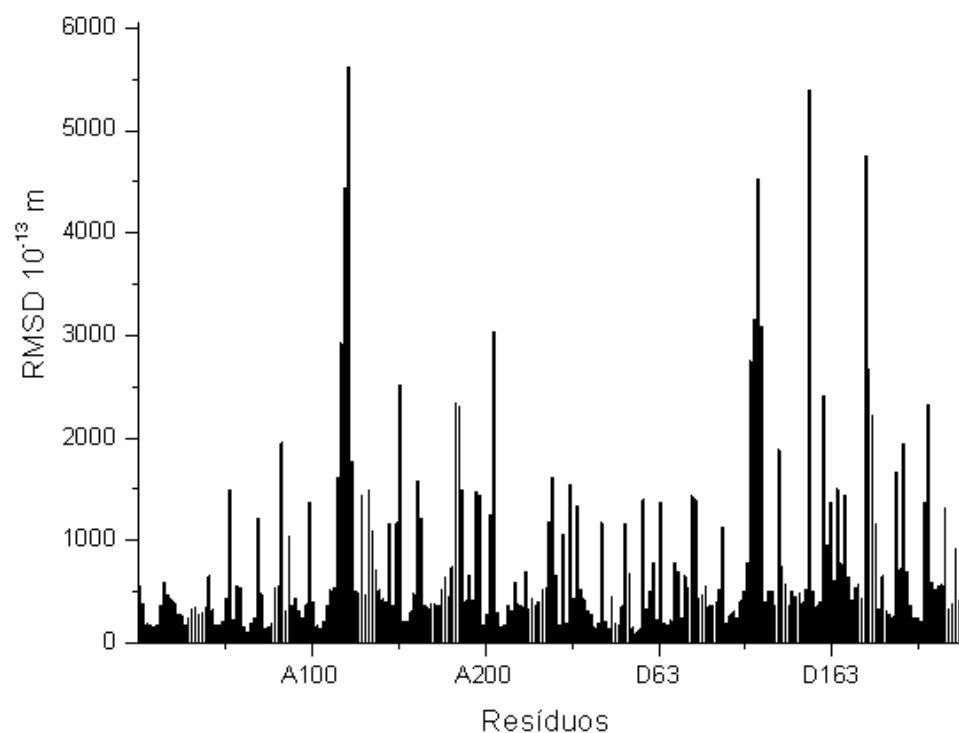


Figura 32.a: Cálculo do RMSD do dímero da lectina CML (calculado para todos os átomos da estrutura) utilizando o programa Profit. Gráfico gerado pelo programa Origin.

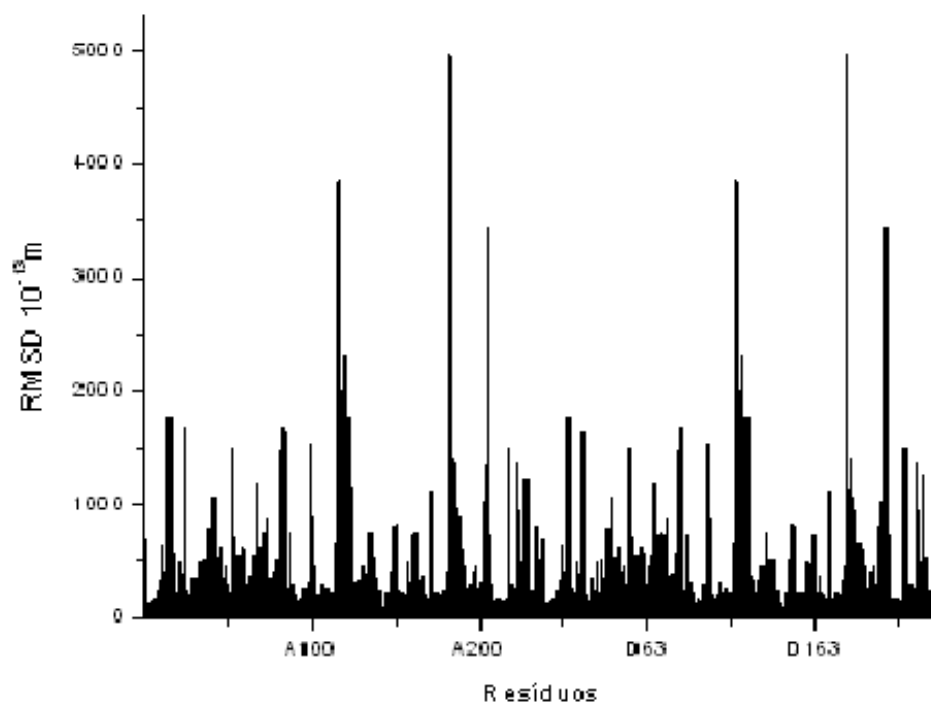


Figura 32.b: Cálculo do RMSD do dímero da lectina CGL (calculado para todos os átomos da estrutura) utilizando o programa Profit. Gráfico gerado pelo programa Origin.

Em alguns pontos na superfície da proteína CML foram encontrados valores relativamente altos para o RMSD individual de alguns resíduos, tais como: Thr¹²⁰, His¹²¹, Thr¹⁵⁰ e Trp¹⁸² com RMSD de 4,9Å, 5,6Å, 4,7Å e 4,5Å, respectivamente. Para a proteína CGL também existem alguns valores relativamente altos para o RMSD individual, tais como: Lys¹¹⁶, Ser¹¹⁹, Thr¹²⁰ e Trp¹⁸² com RMSD de 3,8Å, 2Å, 2,3Å e 4,9Å, respectivamente.

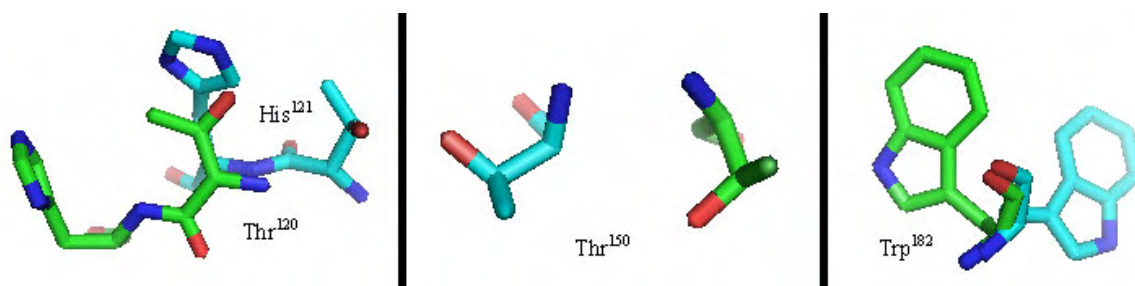


Figura 33.a: Sobreposição dos resíduos com o valor de RMSD relativamente altos da proteína CML nativa (em verde claro) e CML complexada com o açúcar man 1-6 man (em azul claro).

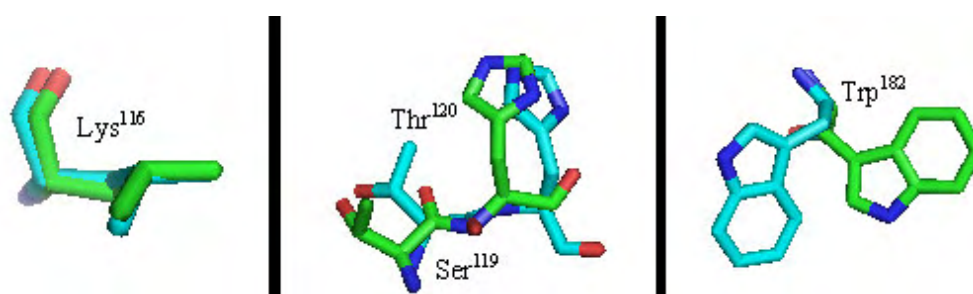


Figura 33.b: Sobreposição dos resíduos com o valor de RMSD relativamente altos da proteína CGL nativa (em verde claro) e CGL complexada com o açúcar man 1-6 man (em azul claro).

Realizou-se o cálculo do RMSD dos átomos da cadeia principal (C, O e N) onde foi calculado o RMSD médio, de 0,841Å (dímero) para a proteína CML e 0,457 Å (dímero) para a proteína CGL. Através do programa Profit foi construída uma tabela e plotado o gráfico desses valores no programa Origin.

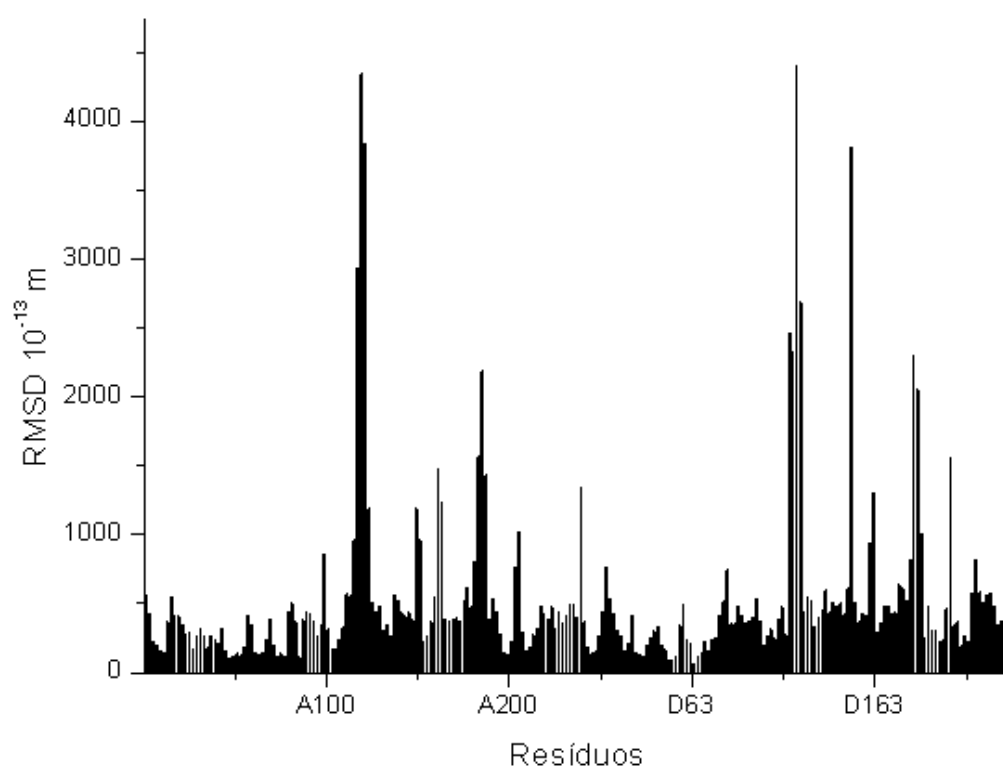


Figura 34.a: Cálculo do RMSD do dímero da lectina CML (calculado para os átomos da cadeia principal) utilizando o programa Profit. Gráfico gerado pelo programa Origin.

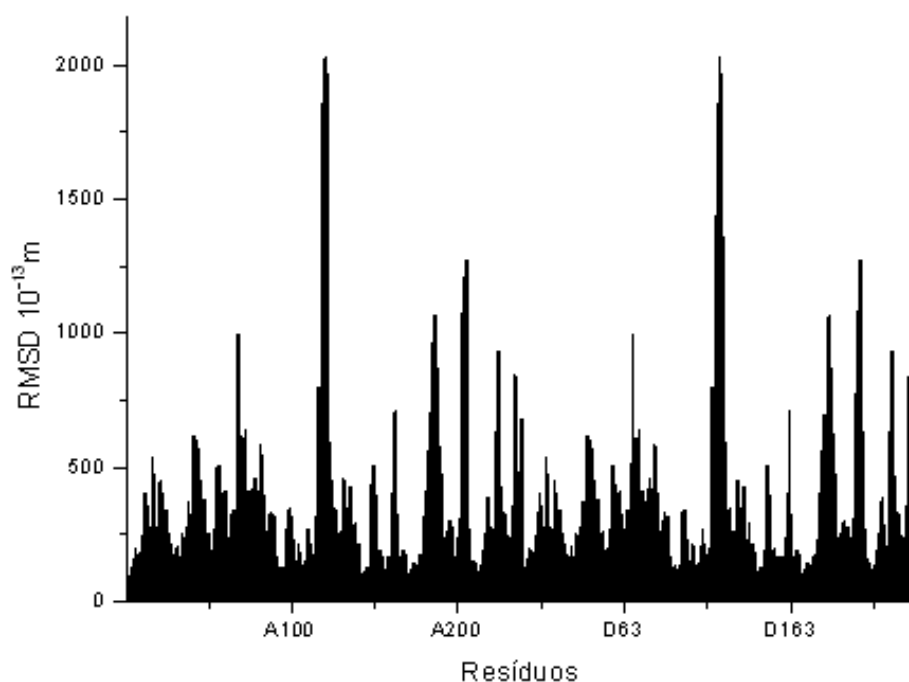


Figura 34.b: Cálculo do RMSD do dímero da lectina CGL (calculado para os átomos da cadeia principal) utilizando o programa Profit. Gráfico gerado pelo programa Origin.

Através dos gráficos podemos observar os picos de RMSD e verificar que para alguns resíduos esses números chegam a $4,5\text{Å}$ (Ser^{119} e Thr^{120} presentes nos dois monômeros), seguidos dos resíduos His^{121} e Thr^{150} que apresentam um RMSD de aproximadamente $3,8\text{Å}$ para a proteína CML, e RMSD de 2Å (Thr^{120} e His^{121} presentes nos dois monômeros) seguidos dos resíduos Ser^{204} e His^{205} que apresentam um RMSD de aproximadamente $1,2\text{Å}$ para a proteína CGL. Além dos resíduos supracitados podemos observar nos gráficos a presença de outros resíduos que também apresentam RMSD significativamente alto, e que influenciam na análise da estrutura quaternária da proteína.

Ajustando somente os resíduos do sítio de ligação, um RMSD de $0,3\text{Å}$ foi observado para a proteína CML mostrando que o sítio catalítico não sofre mudanças conformacionais importantes. Para a lectina CGL, um RMSD médio de $0,2\text{Å}$ foi observado, também, mostrando que o sítio catalítico não sofre mudanças

conformacionais importantes. Corroborando com a idéia de que as mudanças se devem aos diferentes empacotamentos cristalográficos foi feito uma checagem visual, onde se comprovou que os resíduos em questão encontram-se em regiões de contato no empacotamento cristalográfico.

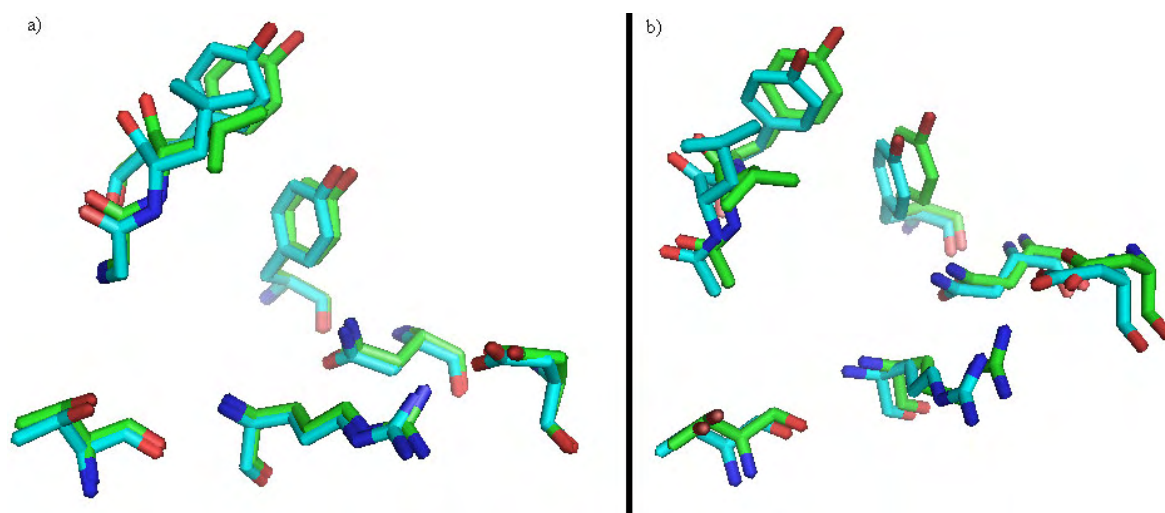


Figura 35.a: Sobreposição dos resíduos do sítio ativo das proteínas CML nativa (em verde claro) e CML complexada com o açúcar man 1-6 man (em azul claro). a) Monômero A. b) Monômero B. Programa utilizado Pymol (DELANO, 2004).

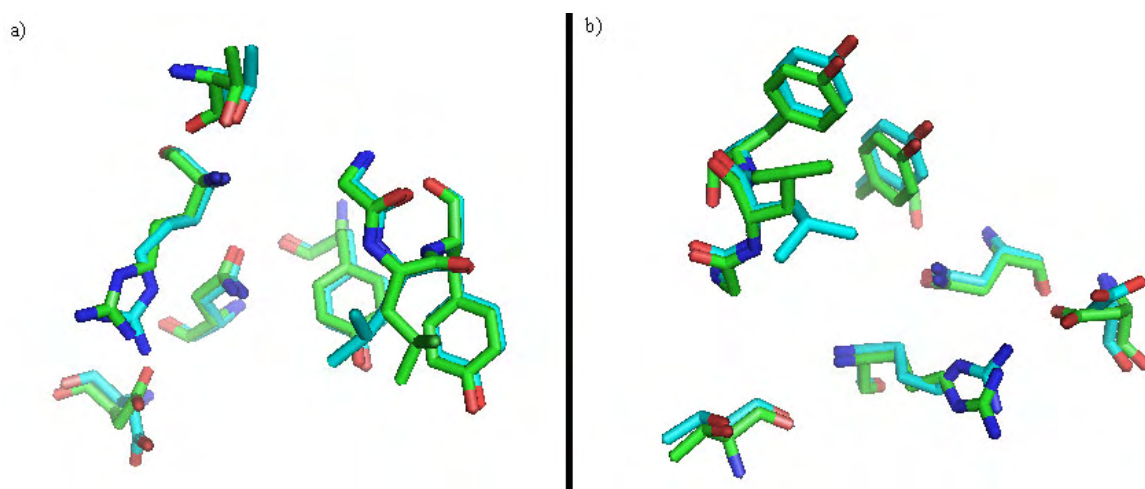


Figura 35.b: Sobreposição dos resíduos do sítio ativo das proteínas CGL nativa (em verde claro) e CGL complexada com o açúcar man 1-6 man (em azul claro). a) Monômero A. b) Monômero B. Programa utilizado Pymol (DELANO, 2004).

4.6 Comparações entre os sítios de ligação

Segue abaixo as descrições e a análise dos sítios de ligação de seis estruturas já resolvidas e depositadas no PDB, sendo três estruturas referentes à proteína CML e três estruturas referentes à proteína CGL:

- CML complexada ao açúcar manose 1-2 manose – OMe (PDB código 2ow4), estrutura tridimensional apresentada na forma monomérica e grupo espacial I222;

- CML complexada ao açúcar manose 1-3 manose – OMe (PDB código 2p37), estrutura tridimensional apresentada na forma tetramérica e grupo espacial P3₂;

- CML complexada ao açúcar manose 1-4 manose – OMe (PDB código 2p34), estrutura tridimensional apresentada na forma tetramérica e grupo espacial P3₂;

- CGL complexada ao açúcar manose 1-2 manose – OMe (PDB código 2ovu), estrutura tridimensional apresentada na forma monomérica e grupo espacial I222;

- CGL complexada ao açúcar manose 1-3 manose – OMe (PDB código 2ef6), estrutura tridimensional apresentada na forma tetramérica e grupo espacial P3₂;

- CGL complexada ao açúcar manose 1-4 manose – OMe (PDB código 2p2k), estrutura tridimensional apresentada na forma tetramérica e grupo espacial P3₂.

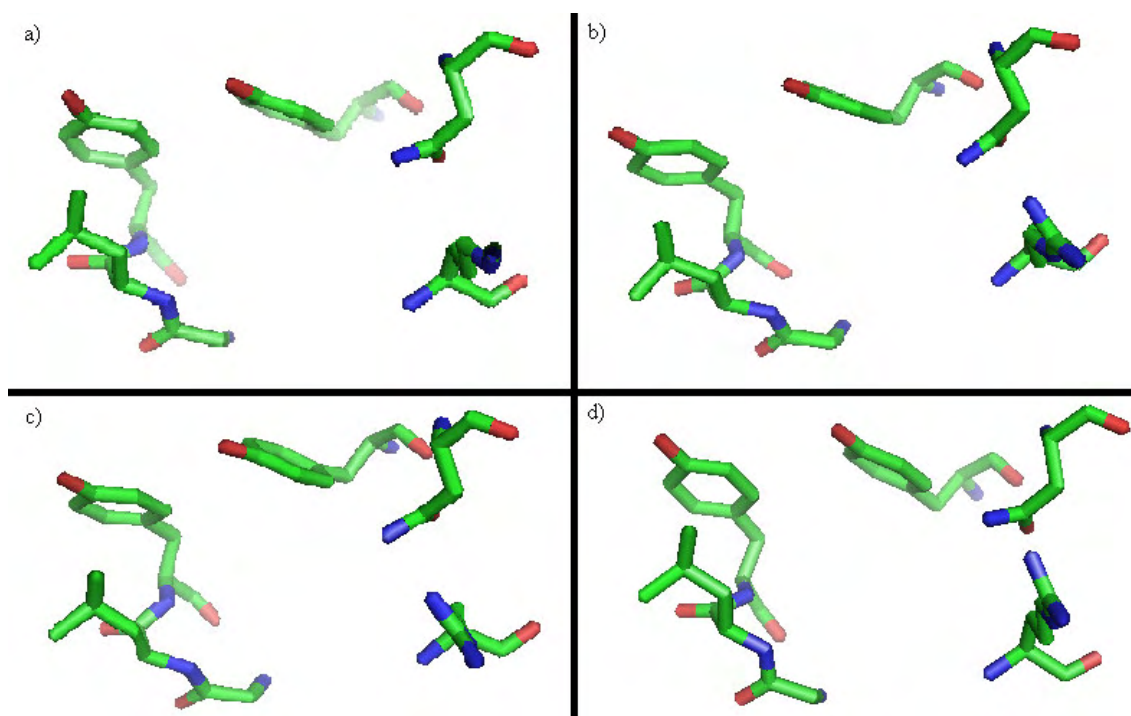


Figura 36: Comparação entre os sítios de ligação da lectina CGL complexada aos açúcares: a) man 1-2 man, b) man 1-3 man, c) man 1-4 man e d) man 1-6 man.

Embora os grupos espaciais das proteínas supracitadas sejam diferentes, observou-se a similaridade entre os sítios de ligação destas com as proteínas CML e CGL complexada ao açúcar manose 1-6 manose. Os resíduos Tyr¹², Asn¹⁴, Gly⁹⁸, Leu⁹⁹, Tyr¹⁰⁰ e Arg²²⁸ estão presentes em todos os sítios de ligação. Outro resíduo presente em todos os sítios de ligação é o Asp²⁰⁸, que estabelece uma interação eletrostática com o açúcar ligado, devido à sua proximidade do mesmo. As duas únicas exceções aparecem nos sítios de ligação das proteínas CML e CGL complexada com o açúcar man 1-2 man – OM, que possuem em seu sítio ativo, além de todos os resíduos citados, dois outros resíduos, Ser¹⁶⁸ e His²⁰⁵, que também estabelecem interações eletrostáticas com o açúcar ligado.

4.7 Comparação entre a CML e a CGL

Como visto anteriormente, análise das estruturas quaternárias das duas lectinas, mostraram-se muito parecidas, o que era de se esperar levando em conta que o alinhamento apresenta uma alta identidade.

A estrutura primária da CML tem 99% de identidade com a estrutura primária da CGL, existindo apenas três resíduos que se diferenciam em suas estruturas. Portanto pouco se pode acrescentar que seja oriundo da comparação das duas estruturas.

5. Conclusão

A resolução da estrutura das proteínas *Canavalia marítima lectin* (CML) e *Canavalia gladiata lectin* (CGL), complexada ao açúcar manose 1-6 manose, fornece informações que contribuem para o entendimento, em nível atômico, dos processos de ligações e inibição do sítio ativo destas proteínas. Além disso, pode-se comparar o sítio de ligação dessas duas lectinas com outras CML e CGL complexada com os açúcares man 1-2 man, man 1-3 man e man 1-4 man (estruturas já resolvida e existente no PDB) e constatar que os sítios de ligação de todas essas lectinas permanecem semelhantes.

Enfim, esse trabalho contribuiu com a resolução de duas estruturas inéditas a serem depositadas no PDB e de certo modo pode ajudar a futuros pesquisadores a realizarem estudos mais completo sobre essas lectinas.

6. Referências

- AGARWAL, R.C.; A new least-squares refinement technique based on the fast Fourier transform algorithm. **Acta Crystallogr.** A34, 791-809 (1978)
- BLUNDELL, T.L.; & JOHNSON, L.N.; **Protein Crystallography**. London: Academic Press (1976)
- BOUCKAERT, J.; HAMELRYCK, T.; WYNS, L.; LORIS, R.; **Novel structures of plant lectins and their complexes with carbohydrates**. *Curr Opin Struct Biol*, **9**:572-577 (1999)
- BOWIE, J.U.; et al., A method to identify protein sequences that fold into a known three-dimensional structure. **Science**, v.253, p. 437-352 (1987)
- CANDURI, F.; Estudo Estrutural de Proteinases Aspárticas por Cristalografia de Difração de Raios X, UNESP, SP. Dissertação de Doutorado (2001)
- CANDURI, F.; FADEL, V.; BASSO, L.A.; PALMA, M.S.; DE AZEVEDO, W.F.Jr.; **Biochem. Biophys. Res. Commun.** V. 327, p. 646-649 (2005)
- CORPET, F.; Multiple sequence alignment with hierarchical clustering. **Nucleic Acids Res.** 25;16(22): 10881-90 (1980)
- DE AZEVEDO, W.F. JR.; Introdução a biocristalografia com o estudo estruturas da quinase dependente de clínica 2 (CDK2) complexada com inibidores, USP, SP (1997)
- DELANO, W.L.; **The PyMOL Molecular Graphics System**. San Carlos: CA DeLan Scientific (2004)
- DE SOUZA, M.M.; Estudos estruturais da Purina Nucleosídeo Fosforilase do *Mycobacterium tuberculosis* complexada com ligantes, UNESP, SP. Dissertação de Mestrado (2007)

DOS SANTOS, D.M.; Estudos estruturais da Purina Nucleosídeo Fosforilase Humana, UNESP, SP. Dissertação de Mestrado (2004)

DRENTH, J.; Principles of Protein X-ray Crystallography (1998)

DUCRUIX, A.; & GIEGE, R.; Crystallization of nucleic acids and proteins. A practical approach. Oxford, UK: IRL Press (1992)

EDELMAN, GM; WANG, JL: **Binding and Functional Properties of Concanavalin-A and Its Derivatives. 3. Interactions with Indoleacetic-Acid and Other Hydrophobic Ligands.** *J Biol Chem*, **253**:3016-3022. (1978)

ELDER, F.R.; GUREWITSCH, A.M.; LANGMUIR, R.V.; et al.; Radiation from electrons in a synchrotron, **Phys. Rev.** 71:829 (1947)

FADEL, V.; Estrutura cristalográfica da hemoglobina isolada do lobo guará (*Chrysocyon brachiurus*), UNESP, SP (2000)

FISHER, R.A.; “On Mathematical Foundations of Theoretical Statistics”, **Philosophical Transactions of the Royal Society of London**, série A 222, p309-368 (1921)

GIACOVAZZO, C.; MONACO, H.L.; VITERBO, D.; SCORDARI, F.; GILLI, G.; ZANOTTI, G.; CATTI, M.; Fundamentals of Crystallography, **Oxford Science Publications** (1991)

KABASCH, W. & SANDER, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers**, v.22, n.12, p.2577-2637 (1983)

LASKOWSKI, R.A., et al.; PROCHECK – A program to check the stereochemical quality of protein structure, **J. Appl. Cryst.**, v.26, p. 507, p. 283-291 (1993)

LUNIN, V.Y. & URZHUMTSEV A.G.; **Acta Cryst.** A40, 269-277 (1984)

MCREE, D.E.; XtalView/Xfit – A versatile program for manipulating atomic coordinates and electron density. **J. Struct. Biol.** 125(2-3):156-65 (1999)

MOCELLIN, A.; Espectroscopia de multico incidência aplicada ao estudo de moléculas, Universidade Estadual de Campinas, Instituto de Física Gleb Wataghin, Laboratório Nacional de Luz Síncrotron (2002)

MORENO, F.B.M.B.; DELATORRE, P.; FREITAS, B.T.; ROCHA, B.A.M.; SOUZA, E.P.; FAÇO, F.; CANDURI, F.; CARDOSO, A.L.H.; FREIRE, V.N.; LIMA, J.L.; SAMPAIO, A.H.; CALVETE, J.J.; DE AZEVEDO, W.F.; CAVADA, B.S.; **Crystallization and preliminary X-ray diffraction analysis of the lectin from *Canavalia gladiata* seeds. *Acta Cryst D*, 60:1493-1495 (2004)**

MURSHUDOV, G.N.; VAGIN, A.A.; DODSON, E.J.; Refinement of macromolecular structures by the maximum-likelihood method. **Acta Crystallogr. D Biol. Crystallogr.** 1;53(pt 3):240-55 (1997)

NAVAZA, J.; AMoRe: an automated package for molecular replacement, **Acta Crystallogr.** A50 157-163 (1994)

NOLASCO, D.O.; Estrutura Cristalográfica da Purina Nucleosídeo Fosforilase do *Mycobacterium tuberculosis*, UNESP, SP. Dissertação de Mestrado (2005)

PEUMANS, W.J.; VAN DAMME, E.J.; **Lectins as plant defense proteins. *Plant Physiol*, 109:347-52 (1998)**

POLIKARPOV, I.; OLIVA, G.; CASTELLANO, E.E.; et al., **Nucl. Instrum. Methods A**, 405, 159-164 (1998)

POLIKARPOV, I.; PERLES, L.A.; DE OLIVEIRA, R.T.; ET AL., Set-up and experimental parameters of the protein crystallography beamline at the Brazilian national synchrotron laboratory **J. Synchrotron Rad.** 5, 72-76 (1998)

RHODES, G.; Crystallography Made Crystal Clean, p. 10-11, Academic Press (1999)

ROSENTHAL, G.A.; **The biochemical basis for the deleterious effects of L-canavanine.** *Phytochemistry*, **30**:1055-1058 (1991)

RUSSO, A.A.; JEFFREY, P.D.; PATTEN, A.K.; et al., **Nature**, (London) 382, 325-331 (1996)

STOUT, G.H. & JENSEN, L.H.; “X-ray Structure Determination”, 2nd ed. Wiley, New York, USA. (1989)

STUART, A. & ORD, J.K.; Classical Inference and relationship, Vol. 2. Kendall’s advanced theory of statistics (5th ed.). New York, Edward Arnald (1991)

SZUHAJ, B.F.; LIST, G.R.; Lecithins, **American Oil Chemists’ Society** (1997)

VAN DAMME, E.J.M.; PEUMANS, W.J.; BARRE, A.; ROUGÉ, P.; **Plant Lectins: A Composite of Several Distinct Families of Structurally and Evolutionary Related Proteins with Diverse.** *Crit Rev Plant Sci*, **17**:575-692. (1998)

VOET, J.G.; VOET, D.; Biochemistry and Molecular Biology Education (BAMBEEd). **Biochem. Educ.** 1;28(3):124 (2000)