



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR
BACHARELADO EM BIOTECNOLOGIA**

DAVID DE ALENCAR ARARIPE

**ESTUDOS ESTRUTURAIS DE DUAS LECTINAS DO GÊNERO *ARACHIS* POR
FERRAMENTAS DE BIOINFORMÁTICA**

FORTALEZA

2017

DAVID DE ALENCAR ARARIPE

**ESTUDOS ESTRUTURAIS DE DUAS LECTINAS DO GÊNERO *ARACHIS* POR
FERRAMENTAS DE BIOINFORMÁTICA**

Monografia apresentada ao Curso de Bacharelado em Biotecnologia do Departamento de Bioquímica e Biologia Molecular da Universidade Federal do Ceará, como requisito à obtenção do título de Bacharel em Biotecnologia.

Orientador(a): Prof^a. Dr^a. Kyria Santiago do Nascimento

FORTALEZA

2017

DAVID DE ALENCAR ARARIPE

**ESTUDOS ESTRUTURAIS DE DUAS LECTINAS DO GÊNERO *ARACHIS* POR
FERRAMENTAS DE BIOINFORMÁTICA**

Monografia apresentada ao Curso de Bacharelado em Biotecnologia do Departamento de Bioquímica e Biologia Molecular da Universidade Federal do Ceará, como requisito à obtenção do título de Bacharel em Biotecnologia.

Aprovada em: 24/11/2017.

BANCA EXAMINADORA

Prof^a. Dr^a. Kyria Santiago do Nascimento (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Benildo Sousa Cavada (Co-Orientador)
Universidade Federal do Ceará (UFC)

Me. Vanir Reis Pinto Júnior
Universidade Federal do Ceará (UFC)

Me. Vinícius José da Silva Osterne
Universidade Federal do Ceará (UFC)

Above all, don't fear difficult moments. The
best comes from them." - Rita Levi-Montalcini

AGRADECIMENTOS

Agradeço primeiramente a Deus pela Vida que nos deu e por toda Criação.

Aos meus pais, Flaminio e Dulcinea, por me trazerem ao mundo, por tudo que me ensinaram e por sempre apoiarem e acreditarem no meu desenvolvimento. Essa conquista também é de vocês! Sou grato também aos meus irmãos Rodrigo e Ana por também terem auxiliado na minha educação e pelo exemplo que transmitiram. Alegria fazer parte desta família, vocês são especiais para mim!

Aos meus professores e orientadores Kyria e Benildo por terem aberto as portas do BioMol-Lab e por terem me acolhido durante a iniciação científica. Sou grato aos senhores pelas oportunidades que me foram dadas e por toda dedicação em manter o grupo forte e unido.

Aos estudantes do BioMol que pude acompanhar durante a iniciação científica: Thatyane, Gleiciane, Luciana, Simoni, Mayara Torquato, Vanir e Vinicius.

A todo grupo do BioMol-Lab pelos momentos de descontração e pelas amizades. Aos companheiros de iniciação científica no grupo: Daniel Lucena, Carol Xavier, Larissa Chicas, Cecília, William, Sara, Hadson, Marliane, Ariany, Lucas e Katiane. Aos amigos da molecular: Luciana, Ivanice, Simoni, Mayara T, Daniel, Cláudia e Cleane. Aos amigos Vanir, Vinicius e Mayara Q., e aos da química de proteínas: Alfa, Júnior, Alysson, Neto e Messias.

Quero também agradecer de forma especial aos alunos de doutorado Vanir e Vinicius por terem aceitado o convite para participar da banca e por toda orientação dada durante a realização dos experimentos e durante os momentos de dúvida. Grato pelo auxílio, pela paciência e por todos os conhecimentos que me transmitiram. Desejo que nossa amizade se preserve por muito tempo e que a vida retribua o empenho de vocês de forma generosa!

Também sou grato pelas amizades que fiz durante esses anos de curso e agradeço de forma especial a cada aluno das turmas 4, 5 e 6 pelos momentos de convivência, de estudo para as disciplinas e por terem tornado esses anos da graduação mais leves. Gratidão especial aos amigos Raissa, Augusto, Carol Xavier, Caroline Landim, Jéssica, Ana Luísa e Clara.

Agradeço também ao auxiliar de coordenação Gilmar e a toda equipe de coordenação do curso de Biotecnologia, a professora Cristina e a professora Marjory, por todo auxílio prestado aos discentes, pelo estímulo e desempenho de cada um para construir um curso cada vez melhor.

Registro também minha gratidão também aos meus amigos de fora da faculdade por estarem presentes quando eu precisei, pelos momentos de diversão que me alegram e por todas conversas, conselhos e aprendizados.

“E aprendi que se depende sempre
De tanta, muita, diferente gente
Toda pessoa sempre é as marcas
Das lições diárias de outras tantas pessoas”

(Gonzaguinha)

RESUMO

Lectinas são proteínas capazes de interagir com carboidratos específicos de forma reversível. Pelo fato dos carboidratos serem moléculas envolvidas em processos de reconhecimento e endereçamento, algumas lectinas são capazes de induzir reações celulares ao interagir com esses açúcares. O gênero *Arachis* pertence à tribo Dalbergieae, um grupo de plantas que tem revelado possuir lectinas promissoras para estudos de atividades biológicas e aplicações biotecnológicas. Este gênero possui três lectinas já purificadas, sendo as três isoladas da mesma espécie, *Arachis hypogaea* (amendoim). Os genomas de duas espécies estreitamente relacionadas ao amendoim comercial, *Arachis duranensis* e *Arachis ipaensis*, foram recentemente sequenciados e publicados. Portanto, com base na informação dos genomas destas duas espécies, este estudo tem como objetivo prever a estrutura tridimensional das lectinas de *Arachis duranensis* (ADL) e de *Arachis ipaensis* (AIL) por modelagem por homologia, analisar a possibilidade de interação desses modelos com diferentes manosídeos através de *docking* (ancoragem; atracamento) molecular e estudar seus comportamentos moleculares através de dinâmica molecular (DM). A conformação tridimensional obtidas para os modelos moleculares está de acordo com os dados obtidos experimentalmente para lectinas de leguminosas estudadas anteriormente, apresentando o motivo *jellyroll* conservado. As duas lectinas tiveram seus domínios de ligação a metais e de reconhecimento à carboidratos caracterizados por sobreposição com estruturas cristalográficas de lectinas de Dalbergieae e *docking* molecular. As lectinas obtiveram escores de ancoragem diferentes quando testados com monossacarídeos, dissacarídeos e *N*-glicanos ricos em manose, sugerindo que podem interagir com esses ligantes com intensidades diferentes. Simulações de DM foram realizadas com as proteínas nas condições sem ligante, interagindo com D-manose e interagindo com α -metil-D-manosídeo. A análise por DM demonstrou aspectos moleculares determinantes para especificidade por α -metil-D-manosídeo. Já durante a simulação com D-manose, as interações moleculares apresentaram menor estabilidade do que as estabelecidas com α -metil-D-manosídeo. Os dados apresentados nesse estudo podem servir como um guia para novas pesquisas realizadas com lectinas α -metil-D-manosídeo específicas do gênero *Arachis*.

Palavras-chave: *Arachis*; lectinas; modelagem por homologia; *docking* molecular, dinâmica molecular;

ABSTRACT

Lectins are proteins capable of interacting with specific carbohydrates in a reversible manner. Because carbohydrates are molecules involved in recognition and addressing processes, some lectins are able to induce cellular reactions when interacting with these sugars. The *Arachis* genus belongs to the Dalbergieae tribe, a group of plants that has shown promising lectins for studying biological activities and biotechnological applications. This genus has already three purified lectins, the three being isolated from the same species, *Arachis hypogaea* (peanut). Genomes of two species closely related to commercial peanuts, *Arachis duranensis* and *Arachis ipaensis*, were recently sequenced and published. Therefore, based on the information of the genomes of these two species, this study aims to predict the three-dimensional structure of *Arachis duranensis* (ADL) and *Arachis ipaensis* (AIL) lectins by homology modeling, to analyze the possibility of interaction of these models with different mannosides through molecular docking and study their molecular behaviors through molecular dynamics (DM). The three-dimensional conformation obtained for the molecular models is in agreement with experimentally obtained data of previously studied leguminous lectins, presenting the conserved jellyroll motif. The two lectins had their metal binding and carbohydrate recognition domains characterized by overlapping crystallographic structures of Dalbergieae lectins and by molecular docking. Lectins obtained different anchorage scores when tested with monosaccharides, disaccharides, and high-mannose N-glycans, suggesting that they may interact with these binders at different intensities. Simulations of DM were performed with the proteins under conditions without binder, interacting with D-mannose and interacting with α -methyl-D-mannoside. DM analysis demonstrated molecular determinants for α -methyl-D-mannosid specificity. During D-mannose simulation, the molecular interactions presented lower stability than those established with α -methyl-D-mannoside. The data presented in this study can thus serve as a guide for new researches aiming α -methyl-D-mannoside specific lectins of the *Arachis* genus.

Keywords: *Arachis* lectins; Molecular Dynamics; Homology modelling

LISTA DE FIGURAS

Figura 1 – Os quatro tipos de lectinas de leguminosas.....	20
Figura 2 –Tipos de arranjos quaternários das lectinas de leguminosas	21
Figura 3 – Relação filogenética entre as tribos de Papilionoideae	24
Figura 4 – Crescimento anual e geral de proteínas depositadas no PDB	31
Figura 5 – Representações estruturais dos ligante MAN3, MAN5 e MAN9	43
Figura 6 – Alinhamento das pré lectinas identificadas contra CTL	46
Figura 7 – Alinhamento das sequências maduras mostrando conservações	47
Figura 8 – Sobreposição das lectinas ADL e AIL com CTL e SL-1	49
Figura 9 – Representação do sítio de ligação a metais conservado para CTL, ADL e AIL.....	50
Figura 10 – Desvios quadráticos médios de todas simulações.....	53
Figura 11 – Flutuações quadráticas médias dos aminoácidos em cada simulação.....	54
Figura 12 – Ligações de hidrogênio estabelecidas entre ADL e seus ligantes.....	55
Figura 13 – Ligações de hidrogênio estabelecidas entre AIL e seus ligantes	56
Figura 14 – Representações capturadas durante simulação da AIL com MAN.....	58
Figura 15 – Representações capturadas durante simulação da AIL com MMA	59
Figura 16 – Representações capturadas durante simulação da ADL com MAN	60
Figura 17 – Representações capturadas durante simulação da ADL com MMA.....	61
Figura 18 – Comparação dos RMSF ADL/AIL + MAN e ADL/AIL + MMA.....	64

LISTA DE TABELAS

Tabela 1 – Lectinas já isoladas da tribo Dalbergieae	24
Tabela 2 – RMSD C α das estruturas de ADL e AIL superpostas com as estruturas SL-1 e CTL. 48	
<i>Tabela 3– Escore do docking da ADL e da AIL</i> obtidos para os diferentes manosídeos	51
Tabela 4– Resíduos de aminoácidos das proteínas ADL e AIL que interagem com MAN e MMA durante as simulações de dinâmica molecular.....	65

LISTA DE ABREVIATURAS E SIGLAS

AAL	<i>Andira anthelmia lectin</i>
ACO	<i>Ant Colony Optimization</i>
ADL	<i>Arachis duranensis lectin</i>
AIL	<i>Arachis ipaensis lectin</i>
ATB	<i>Automated Topology Builder</i>
BanLec	<i>Banana lectin</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
CaBo	Lectina <i>Canavalia bonariensis</i>
ConA	Concanavalina A
ConBr	Lectina de <i>Canavalia brasiliensis</i>
CRD	<i>Carbohydrate binding domain</i> (Domínio de ligação a carboidratos)
CTL	<i>Centrolobium tomentosum lectin</i>
DM	Dinâmica Molecular
DOPE	<i>Discrete Optimized Protein Energy</i>
ExPASy	<i>Expert Protein Analysis System</i>
Gal	Galactose
GalNAc	<i>N</i> -acetil-D-glicosamina
Glc	Glicose
GlcNAc	<i>N</i> -acetil-D-galactosamina
GROMACS	<i>Groningen Machine for Chemical Simulations</i>
LAC	Lactose
LAL	<i>Lonchocarpus araripensis lectin</i>
LBL	<i>Lima bean lectin</i>
LINCS	<i>Linear Constraint Solver</i>
L-NAME	<i>N</i> -nitro-arginina-metil-éster
LSL	<i>Lonchocarpus sericeus lectin</i>
M	Molar
MAL	<i>Machaerium acutifolium lectin</i>
MAN	Manose
MBS	<i>Metal binding site</i> (Sítio de ligação a metais)
MMA	α -Metil-D-manosídeo
PAL	<i>Pterocarpus angolensis lectin</i>

PDB	<i>Protein databank</i>
PELa	<i>Platypodim elegans lectin</i>
PFL	<i>Platymiscium floribundum lectin</i>
PHA-L	<i>Phaseolus vulgaris leucoagglutinin</i>
PI	<i>Isoelectric point</i>
PME	<i>Particle-Mesh Edwald</i>
PNA	<i>Peanut agglutinin (aglutinina do amendoim)</i>
PPL-2	<i>Parkia platycephala lectin 2</i>
QMEAN	<i>Qualitative Model Energy ANalysis</i>
RMSD	<i>Root mean square deviation</i>
RMSF	<i>Root mean square fluctuation</i>
SBA	<i>Soybean agglutinin</i>
SL-1	<i>Stem lectin 1</i>
SL-2	<i>Stem lectin 2</i>
SPC	<i>Single pointe charge</i>
VGL	<i>Vatairea guianensis lectin</i>
VMD	<i>Visual molecular dynamics</i>
VML	<i>Vatairea macrocarpa lectin</i>
WGA	<i>Wheat germ agglutinin (Aglutinina do germe do trigo)</i>
α MM	α -Metil-D-manosídeo

SUMÁRIO

1 INTRODUÇÃO.....	15
1.1 Lectinas.....	15
1.2 Estruturas de lectinas de leguminosas.....	17
1.3 Tribo Dalbergieae.....	22
1.4 O Gênero <i>Arachis</i>	27
1.5 Lectinas de <i>Arachis hypogaea</i>	28
1.6 Bioinformática.....	30
1.5.1 Modelagem molecular.....	30
1.5.2 Docking Molecular.....	34
1.5.3 Dinâmica Molecular em GROMACS.....	38
3 MATERIAIS E MÉTODOS.....	41
3.1 Obtenção da sequência, alinhamento e análise.....	41
3.2 Modelagem molecular e validação.....	41
3.3 Docking molecular.....	42
3.4 Dinâmica molecular.....	43
4 RESULTADOS E DISCUSSÃO.....	45
4.1 Alinhamento e análise de sequências.....	45
4.2 Modelagem molecular e validação.....	48
4.3 Docking molecular.....	51
4.4 Dinâmica molecular.....	52
4.4.1 Desvio quadrático médio e flutuação quadrática média.....	52
4.4.2 Análise de interações.....	56
5 CONCLUSÃO.....	66
6 REFERÊNCIAS.....	67
7 ANEXO I.....	84
8 ANEXO II.....	85

1 INTRODUÇÃO

1.1 Lectinas

As lectinas são proteínas conhecidas há muito tempo e seu descobrimento data de 1888, quando Stillmark publicou sua dissertação sobre o que chamou de “ricina, um fermento tóxico das sementes de *Ricinus communis*” (STILLMARK, 1888). Sua descoberta relacionou o efeito tóxico do extrato de sementes da mamona e sua capacidade de aglutinação a um fator proteico. Em 1898, Elfstrand utilizou pela primeira vez o termo hemaglutinina (ELFSTRAND, 1898), palavra criada para a nova classe de proteínas que vinha sendo identificada.

Com o tempo, a atividade tóxica desse novo grupo proteico deixou de estar relacionada à atividade hemaglutinante. Isso aconteceu quando Landsteiner e Raubitschek reportaram, em 1907, a presença de lectinas atóxicas nas sementes de *Phaseolus vulgaris* (feijão), *Pisum sativum* (ervilha), *Lens culinaris* (lentilha) e *Vicia sativa* (VAN DAMME *et al.*, 1998). Após esse registro, outros pesquisadores encontraram proteínas que também possuíam capacidade de aglutinar células e que não apresentavam a mesma toxicidade observada para ricina. Assim, com a revelação de um maior número dessas novas biomoléculas, a atividade tóxica da ricina passou a ser considerada uma exceção às hemaglutininas (VAN DAMME *et al.*, 1998).

Outro marco importante da história das lectinas aconteceu quando, em 1936, Sumner e Howell publicaram a capacidade da Concanavalina A (ConA) de aglutinar eritrócitos, leveduras, algumas espécies bacterianas e de precipitar glicogênio em solução. Nesse mesmo estudo, os dois cientistas também verificaram que o extrato da cana de açúcar, quando colocado em solução na presença de ConA, inibia a precipitação de amido ou a aglutinação de eritrócitos, sendo o primeiro registro sobre a especificidade de lectinas por açúcares (SHARON & LIS, 2004).

A especificidade das lectinas foi novamente vislumbrada por Boyd e Reguera, em 1949, quando publicaram um estudo que consistiu no teste da atividade hemaglutinante do extrato salino de 262 variedades de plantas contra sangues humanos dos tipos A, B e O. Dos extratos testados, 191 não apresentaram atividade hemaglutinante, 46 apresentaram atividade não específica e 25 apresentaram uma especificidade relativa.

Nessa época, o mecanismo que possibilitava as lectinas aglutinar células ainda era desconhecido. Foram Watkins e Morgan que, em 1952, descobriram que a atividade da lectina de *Phaseolus lunatus* contra sangue humano tipo A era melhor inibida com α -N-acetil-D-galactosamina, e que a atividade da lectina de *Lotus tetranoglobus* contra sangue humano do

tipo O era melhor inibida por α -L-fucose. Assim, eles concluíram que a especificidade dessas lectinas aos tipos sanguíneos A e O se devia à presença desses respectivos açúcares no glicocálice dessas hemácias.

Foi apenas em 1954 que C. Boyd e Elizabeth Shapleigh sugeriram o uso do termo lectina. Seu apelo para o uso do novo termo foi que embora essas proteínas se ligassem a certos antígenos, sua origem não imune impedia que essas proteínas se enquadrassem na mesma classe dos anticorpos, que também podiam aglutinar células. Os autores propuseram então o termo lectina, uma adaptação do participio do verbo “*legere*”, advindo do Latim. Mesmo que criada para enfatizar o fato de que as lectinas possuem afinidade específica a determinado tipo sanguíneo, a ausência de um consenso da comunidade científica fez com que os termos “lectina”, “hemaglutinina” e “aglutinina” fossem utilizados como sinônimos, assim persistindo até os dias atuais (VAN DAMME *et al.*, 1998).

Em seguida, a década de 60 trouxe novos conhecimentos para o campo que depois seria conhecido como lectinologia. Logo no início, Nowell e colaboradores (1960) reportaram que as lectinas de *Phaseolus vulgaris* possuem potencial mitogênico quando colocadas em contato com linfócitos humanos. A constatação desse efeito foi um grande choque para a área de imunologia da época, pois ainda acreditava-se que os linfócitos eram células incapazes de se dividir. Outras publicações importantes feitas nesse período foram os estudos de Aub e colaboradores (1963, 1965), que reportaram que a aglutinina do trigo WGA apresentava preferência por células cancerosas. Essas descobertas estimularam outros autores a pesquisar atividades semelhantes para outras lectinas e assim essas biomoléculas foram se tornando cada vez mais estudadas (TAKAHASHI *et al.*, 1967; MATSUMOTO & OSAWA, 1969; GALBRAITH & GOLDSTEIN, 1970).

Durante esse mesmo período, avanços vinham sendo feitos sobre a interação das lectinas com carboidratos. Goldstein e colaboradores (1963, 1965) relataram em seus resultados que a interação da ConA com polissacarídeos era bem específica, promovendo a formação de precipitados somente quando haviam polissacarídeos contendo resíduos de α -D-glucopiranosil ou α -D-manopiranosil nas extremidades não redutoras. Conhecendo essas propriedades, os pesquisadores Agrawal e Goldstein (1965) supuseram que essa mesma proteína pudesse se ligar aos géis do tipo Sephadex, um polímero reticulado de dextrana comercializado para fazer cromatografias de exclusão molecular, e assim purificá-la de forma mais simples. No artigo, os cientistas relatam que o gel Sephadex G-50 é o mais indicado para fazer a purificação da ConA, pois exclui moléculas de alto peso molecular mais rápido quando comparada com os géis G75, G100 e G200. O gel Sephadex G-25 não foi capaz de reter ConA em sua fase estacionária, o

que foi explicado devido seu alto grau de reticulação. Já o G-200 apresentou maior capacidade de reter moléculas da ConA ligadas, o que pode ser justificado pelo seu baixo nível de reticulação, e, portanto, pode ser utilizada como uma forma de concentrar ConA. Na retirada das proteínas retidas no gel foi utilizada uma solução contendo glicose, que é retirada da proteína por passos de diálise realizados após obtenção da proteína pura.

A introdução e popularização do uso de cromatografias de afinidade para purificação de lectinas acelerou o descobrimento de novas biomoléculas e não tardou para que surgissem publicações relatando-as em outras espécies de plantas, de cogumelos e de animais invertebrados (SHARON & LIS, 2004). Com a disponibilidade de protocolos de isolamento eficientes, muitos laboratórios começaram a publicar as sequências de aminoácidos dessas proteínas e também suas estruturas cristalográficas (HARDMAN & AINSWORTH *et al.*, 1972), o que possibilitou a melhor caracterização da interação dessas moléculas com seus açúcares específicos. Ainda que lectinas tenham sido encontradas em todos os reinos dos seres vivos, as de plantas e, mais especificamente, de leguminosas, receberam uma atenção especial devido seu importante papel na composição de dietas humanas e pela sua importância econômica. As lectinas descobertas nessa família desempenharam papel pioneiro na sua aplicação como ferramentas para pesquisa biológica e biomédica (VAN DAMME *et al.*, 1998).

Muitas dessas proteínas estavam sendo purificadas e publicadas, mas ainda não havia uma definição que pudesse caracterizar de forma precisa todas essas biomoléculas. A primeira definição oficial dada para lectinas foi de Goldstein e colaboradores 1980, mas essa definição considerava a capacidade de aglutinar células ou precipitar polissacarídeos, o que excluía as lectinas monoméricas que são incapazes de desempenhar essas atividades. Já a publicação feita por Peumans e Van Damme em 1995 constata que lectinas de plantas são “todas as proteínas de plantas que possuem no mínimo um domínio não catalítico que se liga reversivelmente a mono ou oligossacarídeos específicos”. Essa definição é aceita até hoje, embora muitos autores adicionem o fato de que essas proteínas não se originam do sistema imune.

1.2 Estruturas de lectinas de leguminosas

No final da década 90, um número relativamente grande de lectinas de leguminosas já era conhecido e importantes trabalhos de revisão foram publicados nessa época. No artigo publicado por Van Damme e colaboradores em 1998, os autores dividem essas biomoléculas em diversas famílias, incluindo dados importantes para cada uma. Até aquele momento, cerca de cem lectinas de leguminosas já haviam sido isoladas de setenta grupos taxonômicos

diferentes. A maioria dessas proteínas foram extraídas de sementes e é registrado que elas podem corresponder a cerca de 1-10% do total das proteínas solúveis a partir de extratos de sementes. Em algumas espécies foram encontradas concentrações de até 50%, enquanto que outras continham apenas 0.1%. Foi também relatado que quando feitas extrações de outras partes da planta, pode-se detectar lectinas que não estão presentes nas sementes, indicando que genes codificantes para lectinas diferentes estão presentes na mesma planta e a expressão delas ocorre em órgãos distintos do vegetal (VAN DAMME *et al.* 1995).

Também em 1998, Loris e colaboradores publicaram uma revisão sobre as estruturas das lectinas de leguminosas. Naquela época já se havia depositado estruturas de 9 lectinas de leguminosas no banco de dados de proteínas (*Protein Data Bank*, PDB), sendo elas Concanavalina A (EDELMAN *et al.*, 1972; HARDMAN & AINSWORTH, 1972), PHA-L (HAMELRYCK *et al.*, 1996), as lectinas da ervilha (EINSPAHR *et al.*, 1986), lentilha (LORIS *et al.*, 1993; LORIS *et al.*, 1994), *Lathyrus ochrus* (BOURNE *et al.*, 1990), *Arachis hypogaea* (BANERJEE *et al.*, 1994; 1996), soja, *Erythrina corallodendron* (SHAANAN *et al.*, 1991), lectina IV de *Griffonia simplicifolia* (DELBAERE *et al.*, 1993) e outras duas estruturas das proteínas homólogas às lectinas, a de *Phaseolus vulgaris* (BOMPARD-GILLES *et al.*, 1996) e da arcelina (HAMELRYCK *et al.*, 1996), nas quais não foram detectadas atividade de ligação a carboidratos.

Essas proteínas apresentam semelhança estrutural e compartilham algumas características. Uma delas é o enovelamento dessas moléculas, que possui um arranjo de folhas beta estruturalmente relacionado com o motivo *jellyroll*, também encontrado em proteínas de capas virais (ARGOS *et al.*, 1980; CHELVANAYAGAM *et al.*, 1992). As estruturas dessas proteínas são constituídas duas folhas beta antiparalelas conectadas por loops (*jellyroll*) conferindo estabilidade estrutural a esse motivo. Algumas das lectinas que possuem esse mesmo arranjo possuem capacidade de sustentar sua atividade mesmo depois de aquecidas a 60 °C por 1 hora (PEREIRA-JÚNIOR *et al.*, 2012; SILVA *et al.*, 2012; LOSSIO *et al.*, 2017).

Outro fator comum a essas lectinas é a conservação de aminoácidos em diferentes regiões. A maioria dessas biomoléculas necessita da ligação com dois cátions divalentes para exercer sua atividade, sendo um deles Ca^{2+} e outro metal, podendo variar entre Mn^{2+} , Zn^{2+} , Co^{2+} ou Ni^{2+} , pois esses íons estabelecem interações com o sítio de ligação a metais (MBS, do inglês *metal binding site*). Através de uma molécula de água, esses metais interagem com um resíduo de ácido aspártico (LORIS *et al.*, 1998) e essa interação é responsável pela manutenção de uma ligação *cis*-peptídica entre Ala-Asp, o que mantém o ácido aspártico em posição *cis* e consequentemente torna a proteína apta a exercer sua atividade. Esse aminoácido se encontra

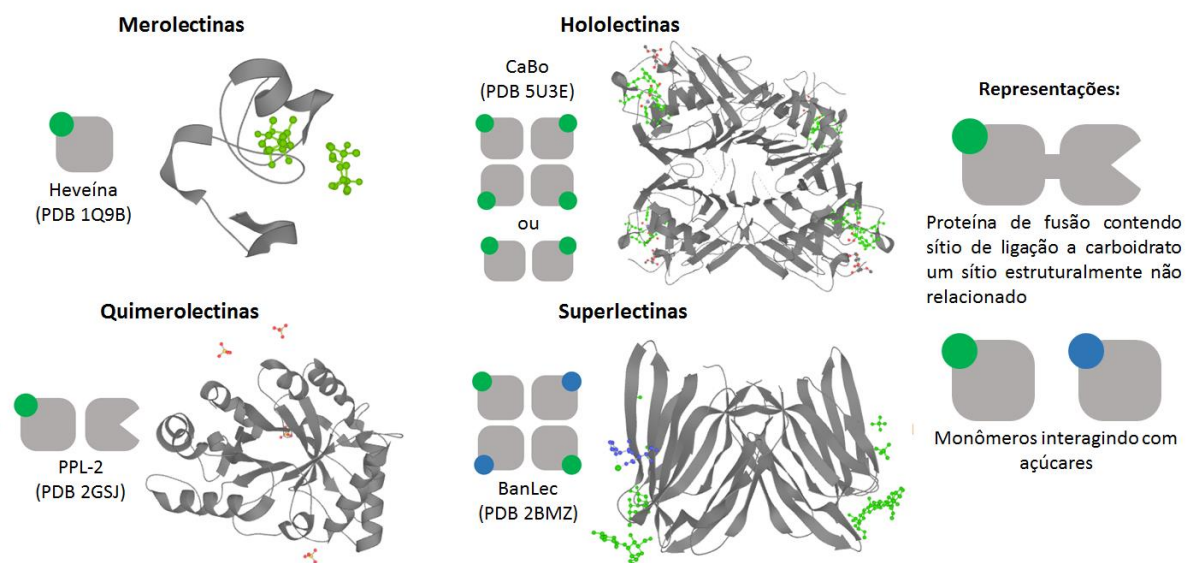
na parte basal do sítio de ligação a carboidratos e estabiliza o ligante dentro do sítio. Tanto os aminoácidos envolvidos no MBS como os aminoácidos envolvidos na isomerização do *cis*-aspartato são altamente conservados nas lectinas de leguminosas, exceto raras exceções (BOMPARD-GILLES *et al.*, 1996; HAMELRYCK *et al.*, 1996). Outros aminoácidos envolvidos no sítio de ligação também são conservados nessas lectinas, mas sua conservação pode variar entre as tribos e subtribos das famílias botânicas em que essas proteínas são encontradas (LORIS *et al.*, 1998).

A subtribo Diocleinae pertencente à tribo Phaseoleae, por exemplo, é uma subtribo que apresenta lectinas com alta homologia entre si (CALVETE *et al.*, 1999) e, apesar dessa homologia, diferentes atividades biológicas são desempenhadas por cada uma dessas lectinas, também conhecidas como *ConA-like* (assim chamadas por sua semelhança à lectina Concanavalina A, isolada de *Canavalia ensiformis*). Devido a sua alta homologia, as lectinas dessa subtribo se tornam bons objetos de estudo para compreensão de como pequenas mudanças na estrutura de uma proteína podem afetar sua atividade biológica. Parte da diferença relatada para as atividades biológicas exercidas por essas lectinas pode ser atribuída às diferentes formas de oligomerização que essas biomoléculas podem assumir, como foi proposto no artigo publicado por Sanz-Aparicio e colaboradores (1997), que fez a comparação estrutural da lectina de *Canavalia brasiliensis* (ConBr) com a lectina de *Canavalia ensiformis* (ConA), que possuem homologia de 99% e, mesmo assim, apresentam diferenças durante ensaios biológicos. As diferentes formas de arranjos diméricos ou tetraméricos podem ser importantes para estabelecer interações biológicas multivalentes, proporcionando maior resistência ao estresse de cisalhamento que está presente em sistemas biológicos de reconhecimento, estabelecendo interações com cinética maior do que sistemas de interação monovalentes e disponibilizando arranjos espaciais diferentes do mesmo sítio de reconhecimento, favorecendo a estabilização de ligações (KIESSLING & POHL., 1996).

Segundo Van Damme e colaboradores (1998), as lectinas podem ser divididas em quatro grupos de acordo com sua estrutura quaternária: merolectinas, hololectinas, quimerolectinas e superlectinas. As merolectinas são aquelas que não se oligomerizam e são encontradas em solução somente na forma de monômeros, como a lectina heveína isolada da espécie *Hevea brasiliensis*. Por conterem apenas um sítio de reconhecimento a carboidratos, essas não manifestam atividade aglutinante e, portanto, são mais difíceis de serem detectadas. Já as hololectinas se agrupam em estruturas quaternárias e possuem sítios de ligação idênticos ou bastante homólogos. Devido sua multivalência, essas proteínas podem aglutinar células e são as mais encontradas dentre lectinas de leguminosas, como exemplos tem-se as lectinas CTL,

PNA, ConBr, ConA, CaBo e diversas outras (ALMEIDA *et al.*, 2016; EINSPAHR *et al.*, 1986; SANZ-APARICIO *et al.*, 1997; HARDMAN & AINSWORTH, 1972). Já as quimerolectinas são proteínas de fusão contendo um ou mais sítios de reconhecimento a carboidratos, possuindo também um sítio estruturalmente não relacionado. Esse outro sítio pode desempenhar atividade enzimática ou outra atividade biológica, mas executa sua função de forma independente do sítio de reconhecimento a carboidratos. Dependendo do seu arranjo espacial, essas proteínas podem se comportar como merolectinas ou hololectinas, e um exemplo desse tipo é a proteína PPL-2 extraída da espécie *Parkia platycephala*, a qual possui um domínio catalítico endoquitinásico e um domínio de reconhecimento a carboidratos específico a *N*-acetil-D-glicosamina. Por último, as superlectinas possuem no mínimo dois sítios de reconhecimento a carboidratos, sendo estes capazes de reconhecer carboidratos estruturalmente não relacionados. Um exemplo desse tipo de lectina é a de banana (BanLec). Todos esses 4 tipos de proteínas de plantas estão representados e exemplificados na Figura 1.

Figura 1 – Representação dos quatro tipos de lectinas de leguminosas..

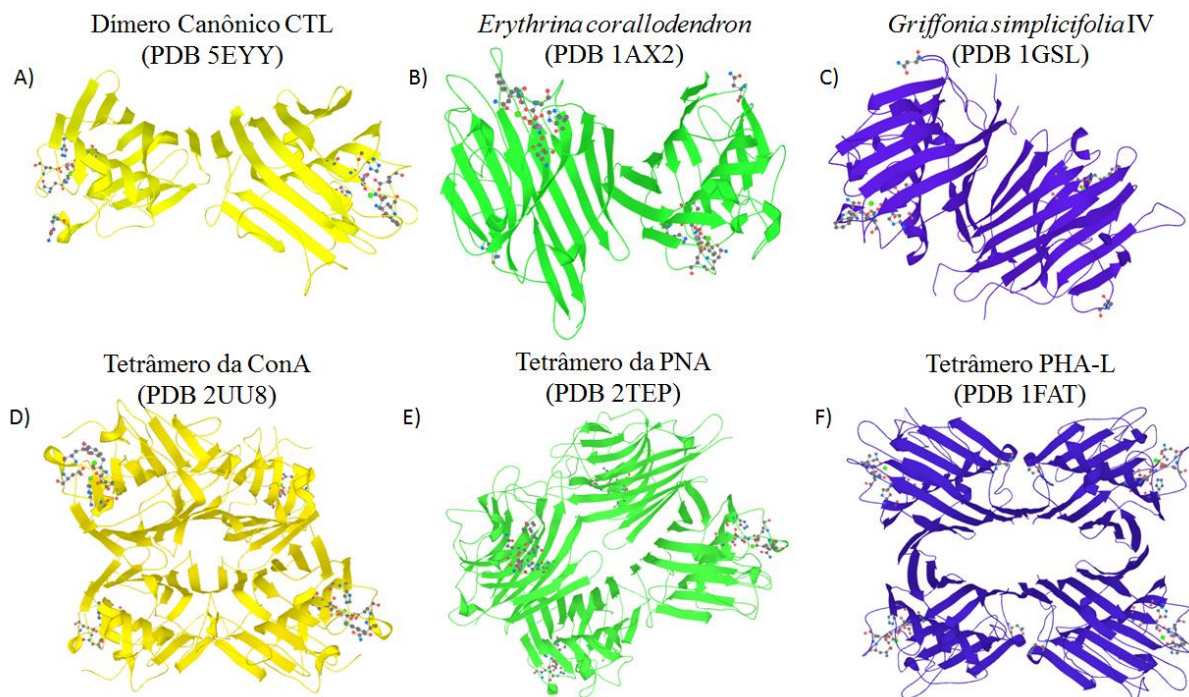


Fonte: Elaborado pelo autor. Adaptado de VAN DAMME *et al.*, 1998

Na família das leguminosas são encontradas hololectinas de diferentes formas. A primeira forma de arranjo é o conhecido dímero canônico das leguminosas, um arranjo descrito primeiramente para lectina de ConA e para a lectina do amendoim (PNA, do inglês *peanut agglutinin*) que consiste em uma extensa folha β composta por 12 fitas resultantes da associação de folhas traseiras de duas lectinas com 6 fitas cada (EDELMAN *et al.*, 1972; HARDMAN e AINSWORTH, 1972, EINSPAHR *et al.*, 1986) (Figura 1 A). Essa associação compreende

moléculas de água que conectam esses monômeros e já foi observado que mutações pontuais nas cadeias de aminoácidos envolvidas nessa interação podem mudar o arranjo das moléculas de água entre elas (LORIS *et al.*, 1998). Outras lectinas de leguminosas apresentam dímeros que interagem de forma diferente do dímero canônico. Um exemplo é a lectina de *Erythrina corallodendron* (Figura 2 B), que supostamente não forma o dímero canônico devido impedimento estérico gerado por um carboidrato ligado covalentemente no resíduo Asn17 (SHAANAN *et al.*, 1991). Outro exemplo é a lectina de lectina IV de *Griffonia simplicifolia* (GS-IV) (Figura 2 C), na qual acredita-se que a presença do aminoácido carregado Glu58 na zona de interação previna a formação do dímero canônico devido repulsão eletrostática (DELBAERE *et al.*, 1993).

Figura 2 – Tipos de arranjos quaternários assumidos pelas hololectinas de leguminosas.



Fonte: Elaborada pelo autor. Adaptado de LORIS *et al.*, 1998

Além dessas interações diméricas, algumas lectinas também são capazes de formar tetrâmeros em solução. O arranjo e comportamento dessas formações são variáveis dentre as lectinas de leguminosas. A lectina ConA, por exemplo, mostrou ser predominantemente tetramérica em pH maior que 7 e predominantemente dimérica em pH menor que 5 (MCCUBBIN & KAY, 1971). Estruturas cristalográficas da ConA e da ConBr revelaram que o tipo de tetrâmero formado por essas proteínas pode sofrer rotação entre seus dímeros (Figura 2 D) e, portanto, supõe-se que esse arranjo apresente comportamento dinâmico em solução

(NAISMITH *et al.*, 1993; LORIS *et al.*, 1996; KANELLOPOULOS *et al.*, 1996; SANZ-APARICIO *et al.*, 1997). Outro arranjo bastante distinto é encontrado na lectina PNA do amendoim (*Arachis hypogaea*), onde há formação de um dímero canônico que interage com outro dímero formado pela interação de loops laterais (Figura 2 E). Estudos anteriores supõem que a formação desse tetrâmero se deve somente às propriedades intrínsecas da molécula, uma vez que essa lectina não é glicosilada e por isso não há impedimento estérico por parte de carboidratos (BANERJEE *et al.*, 1996). Mais uma forma de arranjo foi observada nas lectinas de soja (SBA, do inglês *soybean agglutinin*) e de *Phaseolus vulgaris* (PHA-L), as quais formam tetrâmeros constituídos por dois dímeros canônicos (Figura 2 F) (DESSEN *et al.*, 1995; HAMELRYCK *et al.*, 1996), onde não foi verificada dependência de pH. Esse tipo de tetrâmero foi suposto por Loris e colaboradores (1998), como sendo o padrão para lectinas de leguminosas, por ter sido encontrado com mais frequência do que os formados por ConA e por PNA.

Além dos conhecidos sítios de reconhecimento a carboidratos, outros tetrâmeros de estrutura conhecida também possuem sítios de ligação a moléculas hidrofóbicas (YANG *et al.*, 1974; EDELMAN & WANG, 1978; ROBERTS & GOLDSTEIN, 1983; HAMELRYCK *et al.*, 1999). Ainda são poucos os relatos desses sítios em lectinas, mas normalmente eles estão presentes em proporção de um ou dois para cada tetrâmero (ROBERTS & GOLDSTEIN, 1983; GEGG *et al.*, 1992), sendo alguns deles capazes de estabelecer ligações com hormônios vegetais derivados de adenina (PATHAK *et al.*, 2006; GEGG *et al.*, 1992; MALIARIK *et al.*, 1989).

1.3 Tribo Dalbergieae

A família das leguminosas está dividida em três principais subfamílias, sendo elas Papilionoideae, Cesalpinoideae e Mimosoideae. Dentre elas, a que contém o maior número de lectinas conhecidas é a Papilionoideae, que também é a maior dessas três subfamílias. Nela se distribuem 476 gêneros que contêm 13.860 espécies conhecidas, incluindo a maioria das culturas conhecidas de forragens e de alimentos, bem como espécies utilizadas como modelos genéticos (GEPTS *et al.*, 2005).

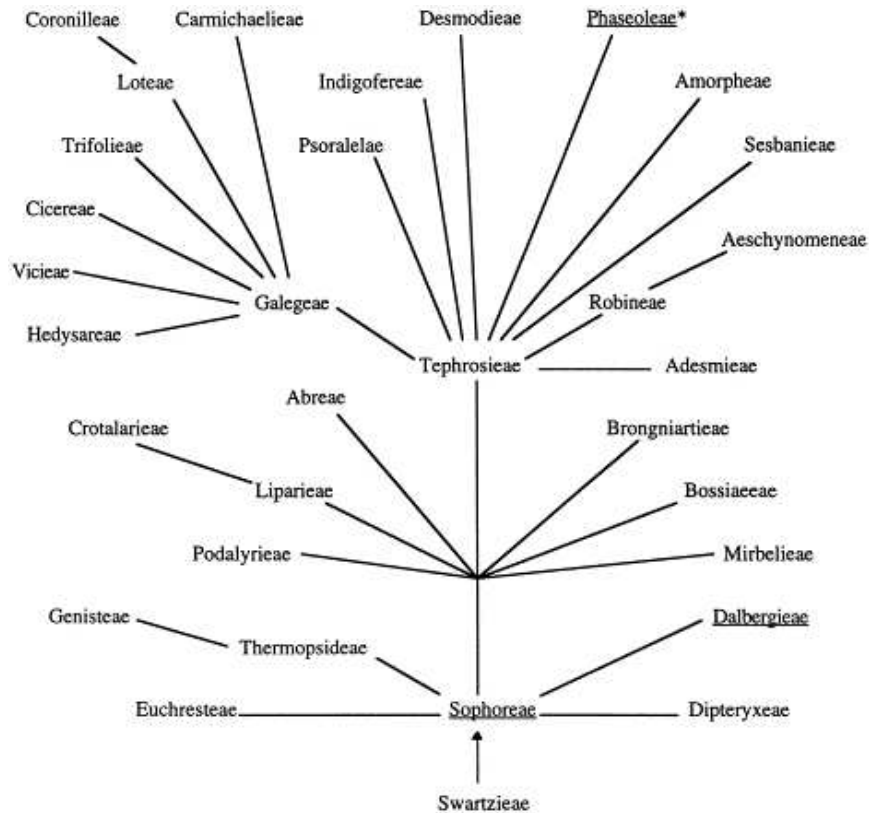
No artigo publicado por Calvete e colaboradores em 1998, o autor propõe a relação filogenética entre as tribos da subfamília Papilionoideae (Figura 3), revelando a posição das Dalbergieae como grupo basal na origem evolutiva. Essa tribo compreende cerca de 49 gêneros e 1324 espécies (LEWIS *et al.*, 2005), das as quais já foram isoladas 16 lectinas.

Além dessa proposta, o autor também propõe que as lectinas pertencentes à tribo

Dalbergieae possuem características peculiares em se tratando de seu processamento pós-traducional. O que foi proposto é que, no processamento pós-traducional da lectina de *Vatairea macrocarpa* (VML), a cadeia α sofre uma clivagem pelo rompimento da ponte entre Asn114 e Lys115. Essa clivagem depende do sítio de glicosilação em Asn111. Após clivada, a cadeia dá origem aos fragmentos γ , com N-terminal deglicosilado e β , com C-terminal glicosilado. Segundo o autor, o fragmento β pode ainda sofrer ainda uma deglicosilação no resíduo Asn183 seguido de processamento N- e C-terminal, gerando isoformas. Quando em solução, esses fragmentos interagem entre si por ligações não covalentes assim manifestam a estrutura tridimensional idêntica à da cadeia alfa (CALVETE *et al.*, 1998).

A presença de isoformas nas lectinas da tribo Dalbergieae é comum (ARARIPE *et al.*, 2017, AGRAWAL *et al.*, 2010) e já foi confirmado que muitas outras lectinas dessa tribo são glicosiladas (ALMEIDA *et al.*, 2016, ARARIPE *et al.*, 2017, NASCIMENTO *et al.*, 2015). Outro efeito que demonstra a presença desse processamento pós traducional é que, quando essas lectinas são submetidas a eletroforese em condição desnaturante, observa-se a presença de diversas bandas no gel (SINGH & DAS, 1994). O mesmo não se observa quando essas lectinas correm a eletroforese em condição nativa ou quando são submetidas a cromatografias de exclusão molecular (PEREIRA-JÚNIOR *et al.*, 2012, ARARIPE *et al.*, 2017), o que indica que os fragmentos gerados pelo processamento estão unidos por interações não covalentes.

Figura 3 – Suposta relação filogenética entre as tribos da subfamília Papilionoideae. Tribos contendo lectinas de duas cadeias estão sublinhadas. * Indica tribo contendo lectinas que sofrem processamento de permutação circular.



Fonte: Calvette *et al.*, 1998

As lectinas já isoladas da tribo Dalbergieae estão listadas na tabela 1. Essas proteínas são bem menos estudadas que as da tribo Phaseoleae. Dentre todas essas lectinas, as mais caracterizadas são as GalNAc/Gal específicas de *Arachis hypogaea* e de *Vatairea macrocarpa* e a Man/Glc específica de *Pterocarpus angolensis*. Algumas atividades biológicas também são descritas para a lectina de *Platypodium elegans*, que apresentou atividade fungicida (BENEVIDES., 2012), proinflamatória (ARARIPE *et al.*, 2017), hipernociceptiva (CAVADA *et al.*, 2017), sendo a primeira das Dalbergieae a apresentar esse efeito dependente do CRD (domínio de ligação a carboidratos; do inglês *carbohydrate binding domain*), e capaz de capturar glicoproteínas em solução.

Tabela 1 – Lectinas já isoladas pertencentes à tribo Dalbergieae

Espécie	Abreviação	Especificidade	Referência
<i>Andira fraxinifolia</i>	--/--	Glc/Man	RANGEL <i>et al.</i> , 2009
<i>Andira pisonis</i>	--/--	Glc/Man	LOSSIO <i>et al.</i> , 2014
<i>Andira surinamensis</i>	--/--	Glc/Man	NOBRE, 2012

<i>Andira anthelmia</i>	AAL	Glc/Man	NASCIMENTO <i>et al.</i> , 2015
<i>Arachis hypogaea</i>	PNA	Gal/GalNAc	LOTAN <i>et al.</i> , 1975
<i>Arachis hypogaea</i>	SL-2	Lac/Celobiose	SINGH e DAS, 1994
<i>Arachis hypogaea</i>	SL-1	α MM	SINGH e DAS, 1994
<i>Centrolobium microchaete</i>	CML	Glc/Man	VASCONCELOS, 2015
<i>Centrolobium tomentosum</i>	CTL	Glc/Man	ALMEIDA <i>et al.</i> , 2015
<i>Lonchocarpus araripensis</i>	LAL	Glc/GlcNAc	PIRES <i>et al.</i> , 2008
<i>Lonchocarpus capassa</i>	--/--	Gal/GalNAc	JOUBERT <i>et al.</i> , 1986
<i>Lonchocarpus sericeus</i>	LSL	Glc/GlcNAc	ALENCAR <i>et al.</i> , 1999
<i>Machaerium acutifolium</i>	MAL	Glc/GlcNAc	BEZERRA <i>et al.</i> , 2003
<i>Platymiscium floribundum</i>	PFL	Glc/Man	PEREIRA-JÚNIOR <i>et al.</i> , 2012
<i>Platypodium elegans</i>	PELa	Glc/Man	BENEVIDES, 2012
<i>Pterocarpus angolensis</i>	PAL	Glc/Man	LORIS <i>et al.</i> , 2003
<i>Pterocarpus rotundifolius</i>	--/--	Glc/Man	MARONDEDZE <i>et al.</i> , 2004
<i>Vatairea guianensis</i>	VGL	Gal/GalNAc	SILVA <i>et al.</i> , 2012
<i>Vatairea macrocarpa</i>	VML	Gal/GalNAc	CAVADA <i>et al.</i> , 1998

Adaptado de Martins, 2017

A lectina de *Lonchocarpus sericeus* também já possui atividade biológica publicada na literatura. O artigo publicado por Napimoga e colaboradores (2007) revela que essa proteína foi capaz de atenuar os processos de migração de leucócitos para o endotélio, de transmigração de leucócitos e da hipernocicepção causada pela injeção de um estímulo inflamatório. Segundo Pires e colaboradores (2008), a lectina de *Lonchocarpus araripensis* (LAL) possui ação anti-inflamatória sobre o edema de pata induzido por carragenina, porém é desprovida de ação pró-inflamatória quando injetada localmente. Essa mesma proteína também demonstrou atenuar a inflamação aguda celular em ratos (PIRES *et al.*, 2016).

A primeira lectina dessa tribo que demonstrou atividade vasorelaxante foi a de *Vatairea guianensis* (SILVA *et al.*, 2012). Antes dela, esse tipo de atividade só havia sido relatada para lectinas da tribo de Diocleinae (ASSREUY *et al.*, 2009; NOBREGA *et al.*, 2012). Essa atividade foi revertida pela adição do açúcar específico de VGL, o que indica que essa atividade está relacionada com uma ação do CRD da proteína. O efeito também foi revertido pela adição de L-NAME, uma molécula inibidora da enzima óxido nítrico sintase (SILVA *et al.*, 2012). Já em ensaios edematogênicos, essa mesma lectina demonstrou agir de forma

independente do óxido nítrico, pois a adição de L-NAME não foi capaz de inibir a formação do edema (MARQUES *et al.*, 2017). Uma hipótese que poderia explicar esse efeito é a expressão de diferentes isoformas da enzima óxido nítrico sintase nos vasos não inflamados e nos tecidos da pata inflamada (SALVEMINI *et al.*, 1996).

Outra lectina da tribo Dalbergieae que apresentou atividade edematogênica foi a da espécie *Centrolobium tomentosum* (ALMEIDA *et al.*, 2016). Os resultados obtidos também revelaram que o efeito edematogênico de CTL foi inibido pela adição de seu açúcar específico. O efeito elicitado pela CTL demonstrou maior intensidade no intervalo de 2 a 4 horas, com pico máximo em 2 horas após injeção (ALMEIDA *et al.*, 2016). Ambas as lectinas VML e CTL apresentaram atividades edematogênicas dependentes do CRD, apesar das suas especificidades diferentes (Tabela 1).

A lectina de *Andira anthelmia* (AAL) apresentou efeito analgésico envolvendo o domínio de lectina por meio de mecanismos periféricos de nocicepção inflamatória (NASCIMENTO *et al.*, 2015).

Das proteínas citadas na tabela 1, apenas as lectinas PNA (BANERJEE *et al.*, 1996), CTL (ALMEIDA *et al.*, 2016), PELa (BENEVIDES *et al.*, 2012; CAVADA *et al.*, 2017), PAL (BANERJEE *et al.*, 1996) e VML (SOUSA *et al.*, 2015) possuem estruturas tridimensionais depositadas no banco de dados PDB (*Protein Data Bank*). Uma vez conhecida a estrutura dessas lectinas, é possível observar suas interações com seus carboidratos específicos, caracterizar o comportamento da proteína em resposta à solução, engenheirar e aprimorar a estrutura por mutação dirigida e realizar triagens virtuais. As possibilidades de aplicação desses dados estruturais são muito vastas e, portanto, a determinação de uma estrutura tridimensional é fundamental para caracterizar e investir em uma biomolécula.

A lectina de *Pterocarpus angolensis*, PAL, já foi cristalizada com alguns ligantes diferentes e no PDB consta trinta estruturas diferentes depositadas para essa mesma proteína, o que a torna a proteína de Dalbergieae com mais estruturas depositadas. Seu sítio de ligação já foi caracterizado em complexo com α -metil-D-glicose, sacarose, turanose (LORIS *et al.*, 2003), α -metil-D-manosídeo, dimanosídeos com ligações glicosídicas do tipo α 1-2, α 1-3, α 1-4 e α 1-6, trimanosídeo Man(α 1-6)Man(α 1-3)Man (LORIS *et al.*, 2004) e com os glicanos ricos em manose MAN6, MAN7D1, MAN7D3, MAN-8D1D3 e MAN9 (GARCIA-PINO *et al.*, 2007). Em outro artigo, a estrutura da PAL foi elucidada na presença do decassacarídeo complexo NA2F e com outros quatro oligossacarídeos que são fragmentos desse decassacarídeo (BUTS *et al.*, 2006). Apesar dos minuciosos estudos estruturais realizados, essa lectina não possui atividade biológica caracterizada na literatura.

A lectina de *Vatairea macrocarpa* possui 6 estruturas depositadas no banco de dados de proteínas com seu domínio interagindo com os ligantes GalNAc, antígeno Tn e lactose. Dentre as atividades biológicas já relatadas para essa lectina, pode-se citar a indução de edema na pata de ratos após injeção intraplantar (ALENCAR *et al.*, 2004). Seu pico de atividade aconteceu após 4 horas do momento de injeção, diferindo do ensaio promovido com a CTL. VML também foi capaz de induzir migração de neutrófilos e células mononucleares para cavidade peritoneal de ratos (ALENCAR *et al.*, 2003). Nesse trabalho, propõe-se que essa lectina induz a migração de neutrófilos por uma via indireta mediada por mastócitos, pois a diminuição de células residentes do tecido peritoneal diminui significativamente a migração de células de defesa e porque a presença atenuada de macrófagos na área de injeção potencializou o efeito da lectina (ALENCAR *et al.*, 2003). Os efeitos verificados demonstraram depender do CRD da lectina, uma vez que a injeção da proteína desnaturada não surtiu efeito e que a incubação prévia com galactose foi capaz de inibir parcialmente o recrutamento de neutrófilos (ALENCAR *et al.*, 2003). Outras das aplicações de VML incluem a indução do efeito depressivo e a expressão de marcadores neuroinflamatórios em ratos (GONÇALVES *et al.*, 2013), o aumento da resistência vascular renal, da taxa de filtração glomerular e do fluxo urinário (MARTINS *et al.*, 2005). Essa lectina também demonstrou especificidade ao antígeno Tn, comumente encontrado em células cancerígenas (DAM *et al.*, 2007).

1.4 O Gênero *Arachis*

O gênero *Arachis* pertence à tribo Dalbergieae e compreende um total de 69 espécies, das quais a mais conhecida é a *Arachis hypogaea*, o amendoim. A atual produção mundial de amendoim por ano é de cerca de 46 milhões de toneladas, sendo a leguminosa mais produzida nos continentes Ásia e África (<http://www.fao.org/faostat/en/#home>) e a noz mais consumida dos Estados Unidos. O consumo dessa semente pode ser feito de forma crua, assada, cozida ou pelos seus derivados como óleo de amendoim, manteiga de amendoim e até amendoim em farinha, que pode servir como alternativa para confecção de pães sem glúten.

Essa planta é endêmica da América do Sul e sua utilização na dieta data da época dos Incas, no Peru. A popularização do amendoim só começou durante a guerra civil dos Estados Unidos durante a década de 60. A invenção da manteiga de amendoim surgiu na década de 90 e foi inventada por um médico de Saint Louis como uma reposição proteica para pessoas com problemas dentários. Outras formas de comer amendoim surgiram do agrônomo George Washington Carver, que lançou um livro intitulado “Como cultivar o amendoim e 105 maneiras

de prepará-lo para o consumo humano” (ARYA *et al.*, 2016).

Já foi relatado que esse alimento contém diversos compostos bioativos preventivos para doenças no coração e outros promovedores da longevidade (ARYA *et al.*, 2016), além de conter todos os 20 aminoácidos. Recentemente, foi revelado que o amendoim é excelente fonte de compostos como resveratrol, ácidos fenólicos, flavonóides e fitoesteróis que bloqueiam a absorção de colesterol da dieta. Também já foi comprovado que seu consumo a longo prazo diminui o risco de obesidade (SABATE & ANG, 2009).

Apesar de todos esses benefícios que o amendoim proporciona à saúde humana quando acrescentado nas dietas, o amendoim também é um substrato propício para fungos que podem produzir aflatoxinas, cuja toxicidade representa uma ameaça séria à saúde (NJAPAU *et al.* 1998; BHAT & VASHANTI 1999).

Esse problema ameaça a produção do amendoim e, portanto, se tem interesse de obter o genoma da *Arachis hypogaea* como uma fonte para pesquisar possíveis melhoramentos genéticos que possam aumentar a resistência dessa espécie contra fungos e outras pragas. Porém, a *A. hypogaea* é uma espécie tetraploide e seu genoma possui cerca de 64% de sequências repetidas (DHILLON *et al.*, 1980), o que dificulta muito o sequenciamento do seu genoma. A fim de fornecer uma maneira alternativa para melhor entender o genoma do amendoim comercial, um estudo recente objetivou sequenciar o genoma (BERTIOLI *et al.*, 2016) de dois espécimes intimamente relacionados, *Arachis duranensis* e *Arachis ipaensis* (NIELEN *et al.*, 2012).

A espécie *A. hypogaea* tem genoma constituído de quatro conjuntos de cromossomos, sendo eles dois conjuntos de cromossomos A e dois conjuntos de cromossomos B. Alguns estudos realizados indicam que *A. duranensis* e *A. ipaensis* correspondem aos doadores desses dois subgenomas do amendoim (KOCHERT *et al.*, 1996, MORETZSOHN *et al.*, 2013, RAMOS *et al.*, 2006, GRABIELE *et al.*, 2012). Os tamanhos dos genomas de cada uma dessas três espécies também estão de acordo com o que já foi suposto, pois o tamanho do genoma de *A. hypogaea* de aproximadamente 2,7 Gb corresponde à soma dos 1,25 Gb de DNA de *A. duranensis* com os 1,56 Gb de DNA de *A. ipaensis* (SAMOLUK *et al.*, 2015).

1.5 Lectinas de *Arachis hypogaea*

Do amendoim já foram isoladas três lectinas: PNA, SL-1 e SL-2. A mais conhecida delas é a PNA, extraída das sementes da planta, que possui especificidade a galactose. Essa lectina é conhecida pelas suas atividades antitumorais, sendo capaz de induzir apoptose e autofagia em células cancerosas por uma ligação seletiva ao antígeno T presente nessas células

(MUKHOPADHYAY *et al.*, 2014).

O banco de dados de proteínas contém 23 estruturas depositadas e a mais antiga delas data de 1986 (EINSPAH *et al.*, 1986). Os dados estruturais obtidos até então revelaram que essa lectina não possui glicosilação, o que pode ser indício de que ela não passe pelo mesmo processamento pós traducional que a maioria das Dalbergieae passam.

Diversas aplicações biotecnológicas já foram relatadas para PNA na literatura como sua utilização para promover separação de células imunoincompetentes do tímócito através da seletividade da lectina (REISNER *et al.*, 1976), para fazer sinergia com anticorpos na detecção de um biomarcador do câncer de mama (PARK *et al.*, 2013), para inibir a proliferação de células de câncer colorretal *in vitro* (LOREA *et al.*, 1997) e para inibir a proliferação de células de melanoma *in vitro* (KISS *et al.*, 1997).

As lectinas SL-1 e SL-2, diferentemente da PNA, foram extraídas das hastes da *Arachis hypogaea* e elas possuem especificidade a α MM e a Lac/Celobiose, respectivamente (SINGH & DAS, 1994). Elas foram purificadas através dos géis cromatográficos Sephadex G-50 (SL-1) e de goma de guar (SL-2), sendo retiradas das matrizes com soluções de glicose 0,2 M (SL-1) e de galactose 0,2 M (SL-2). O teste de quantificação de carboidratos revelou que essas proteínas são glicosiladas e contêm 10% e 40% de açúcares neutros em suas composições, respectivamente. A eletroforese em condição desnaturante apresentou diversas bandas para cada uma dessas proteínas, mas quando o teste foi feito em condições nativas, ambas apresentaram uma única banda (SINGH & DAS, 1994), o que pode indicar que essas lectinas passam por um processamento pós traducional semelhante ao da lectina de VML (CALVETE *et al.*, 1998).

Até hoje, cientistas já sequenciaram, caracterizaram e obtiveram a estrutura de uma das isoformas da SL-1 por modelagem molecular (AGRAWAL *et al.*, 2010). Um outro estudo realizado com essa mesma lectina relata sua expressão em *Escherichia coli*, sua purificação, seu sequenciamento, sua estrutura tridimensional através da técnica de modelagem comparativa e uma atividade antagonista a citocinina promovida pelo sítio de ligação hidrofóbico presente na SL-1 (PATHAK *et al.*, 2006). A suposição da presença do sítio de ligação hidrofóbico nessa lectina foi decorrente da observação do resultado da modelagem molecular e da sua comparação com as lectinas PHA-E e LBL. A SL-1 nesse trabalho não apresentou glicosilação já que o aparato molecular da *E. coli* não permite esse tipo de modificação.

Embora o estudo realizado por Singh e Das (1994) tenha relatado que a SL-2 foi inibida fortemente pelo antígeno T Gal β 1 \rightarrow 3GalNAc, não foram feitos outros estudos com essa lectina. Seu potencial de ligação com esse antígeno pode ser um indício de que essa lectina possa ser aplicada como ferramenta biotecnológica de forma semelhante à PNA.

1.6 Bioinformática

Com o advento de novas tecnologias de sequenciamento, muitas espécies tiveram seus genomas sequenciados, possibilitando o conhecimento de muitas proteínas antes mesmo delas terem sido purificadas. A obtenção de estruturas proteicas ainda hoje depende muito da cristalografia de raios-X, técnica que, embora muito eficiente, possui certas desvantagens como depender da proteína em alto grau de pureza, ter seu processo dificultado caso haja grupos com alto grau de liberdade (loops ou glicosilações extensas) e depender de técnicas de cristalização que são afetadas por diversos fatores.

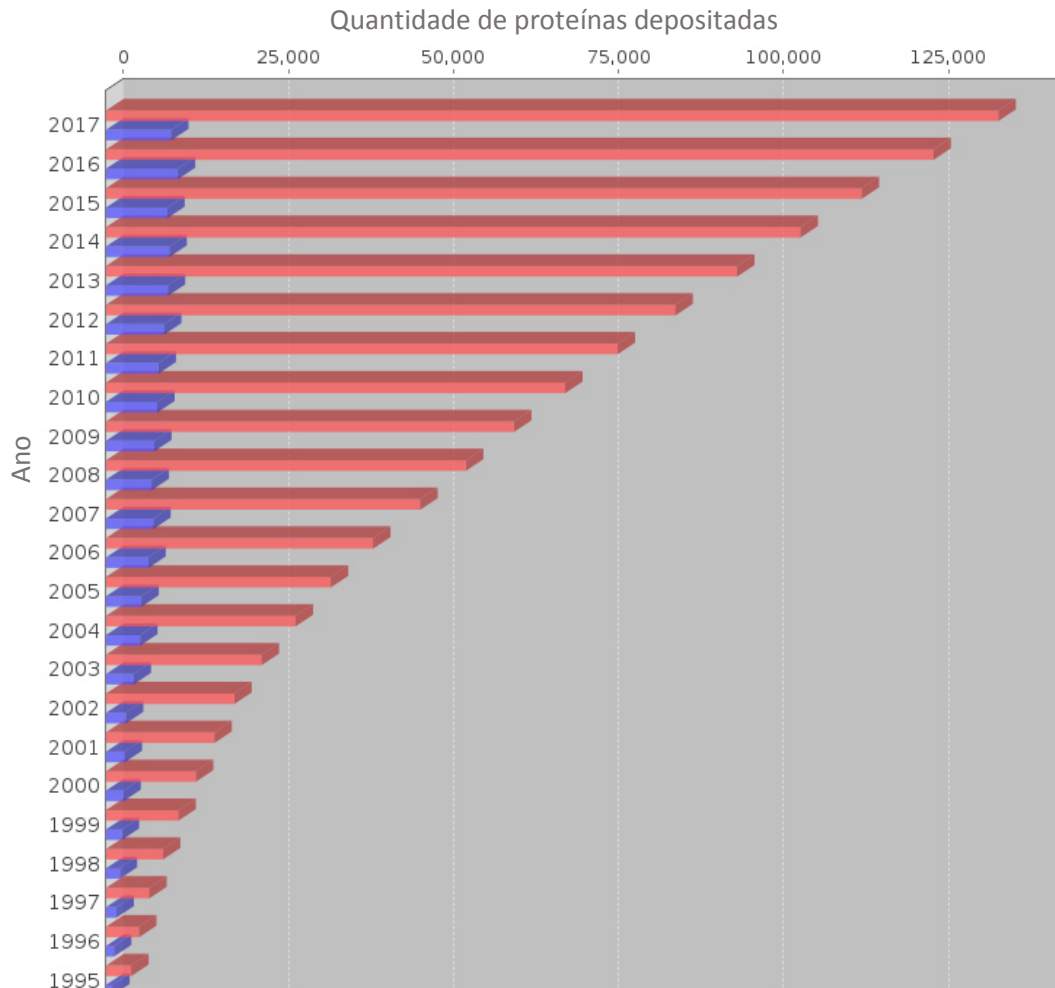
1.5.1 Modelagem molecular

Hoje, com muitas proteínas depositadas no banco de dados, com a disponibilidade de computadores de alto poder de processamento e de linguagens de programação, pode-se obter estruturas tridimensionais de proteínas através do desenvolvimento de diferentes algoritmos. Muitos algoritmos já foram desenvolvidos com esse propósito, mas o método disponível que apresenta maior precisão é o de modelagem por homologia ou modelagem comparativa (KOEHL *et al.*, 1999), que requer a utilização de uma proteína molde cuja sequência de aminoácidos possua no mínimo 25% de homologia com proteína a ser modelada. Embora o dado gerado seja apenas teórico, já é comprovado que a modelagem comparativa é capaz de gerar estruturas com uma acurácia comparável a estruturas experimentais de baixa resolução (MARTÍ-RENOM *et al.*, 2000). Essa característica faz da modelagem molecular uma boa ferramenta para análise de estrutura geral, identificação e caracterização de sítios de ligação e para o planejamento de futuros experimentos (FISER & SALI, 2003).

A eficiência dessa técnica se deve ao fato de que as estruturas terciárias das proteínas são muitas vezes mais conservadas do que suas próprias sequências (LESK *et al.*, 1980). Um exemplo disso são as próprias lectinas de leguminosas que mesmo com mudanças em diversos aminoácidos, apresentam o mesmo motivo *jellyroll* e um arranjo semelhante de loops. Outro fato que faz essa técnica uma ferramenta potencial para estudos estruturais é o desenvolvimento da biologia estrutural desde a publicação da primeira estrutura proteica de alta resolução em 1958 por John Kendrew e Max Perutz. Hoje, o banco de dados de proteínas já contém um total de 125.511 estruturas depositadas (<https://www.rcsb.org/pdb/statistics/holdings.do>; acessado novembro de 2017), sendo a grande maioria delas obtidas pela técnica de cristalografia de raios-X. O número de proteínas

depositadas por ano vem se mostrando aproximadamente constante a 5 anos e pode ser observada na Figura 4.

Figura 4 – Crescimento anual e total de estruturas depositadas no PDB (*Protein Data Bank*). Em vermelho: total de proteínas depositadas. Em azul: proteínas depositadas no ano



Fonte: <https://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100> acessado em novembro de 2017.

Essa grande disponibilidade de estruturas de proteínas permitiu que uma sequência proteica deduzida de forma aleatória dentro de um genoma tivesse 20 a 65% de chance de estar relacionada com outras estruturas depositadas dentro do banco de dados, dependendo do genoma do qual a sequência foi obtida (SANCHEZ *et al.*, 2000, JENNINGS *et al.*, 2001). Assim, com o crescimento constante de proteínas depositadas, a metodologia de modelagem molecular por homologia poderá ser utilizada no futuro para a grande maioria das proteínas.

A aplicação dessa técnica usualmente se dá a partir dos seguintes passos: identificação de estruturas semelhantes, seleção de modelos moleculares, alinhamento de sequência com a estrutura modelo, construção da proteína em estudo utilizando informações da estrutura modelo e validação da estrutura (FISER & SALI, 2003). Afim de se obter uma

estrutura com bons parâmetros de validação, é importante que a estrutura que servirá como modelo tenha boa resolução e um bom fator Rfree (medida entre as amplitudes dos fatores estruturais da proteína gerada em comparação com os fatores calculados para os dados originais de difração; evita processos de super refinamento, onde a estrutura cristalográfica obtida não condiz com os dados experimentais). Por isso, se a estrutura modelo tiver diferentes conjuntos depositados, é preferível que o pesquisador escolha aquela de maior resolução.

Uma vez que a proteína modelo esteja escolhida, diversos métodos podem ser utilizados para gerar a estrutura da proteína que se quer modelar por homologia. Uma delas é a montagem de corpo rígido (BROWNE *et al.*, 1969; GREER, 1981, BLUNDELL *et al.*, 1987), que constrói um modelo a partir de poucas regiões centrais e também de loops e cadeias laterais, que são obtidas pela pesquisa de estruturas relacionadas. Outro método é a modelagem por correspondência de segmentos (JONES & THIRUP 1986; UNGER *et al.*, 1989; CLAESSENS *et al.*, 1989, LEVITT, 1992), que se baseia nas posições aproximadas dos átomos conservados com os moldes para calcular as coordenadas dos outros átomos. A terceira forma de modelagem por homologia é a utilizada pelo programa MODELLER (FISER & SALI, 2003), a qual modela cumprindo restrições espaciais utilizando geometrias de distância ou técnicas de otimização para satisfazer as restrições obtidas do alinhamento da sequência alvo com a estrutura molde (GREER, 1981; HAVEL & SNOW, 1991; SRINIVASAN *et al.*, 2008; BROCKLEHURST & PERHAM, 2008; ASZODI & TAYLOR, 1996).

A ferramenta MODELLER extrai as restrições espaciais de duas fontes. Primeiramente, as restrições de distâncias e ângulos diedros da proteína molde são extraídas e aplicadas na sequência alvo correlacionando com o alinhamento. Em seguida, as restrições estereoquímicas, tais como o comprimento das ligações e as preferências dos ângulos de ligação, são obtidas a partir do campo de força de mecânica molecular CHARMM-22 (MACKERELL *et al.*, 1998), e as preferências estatísticas dos ângulos diédricos e as distâncias atômicas não ligadas são obtidas a partir de um conjunto representativo de todas as estruturas proteicas conhecidas. O modelo é então calculado por um método de otimização com base em gradientes conjugados e em dinâmica molecular, o que minimiza as violações das restrições espaciais (FISER & SALI, 2003).

Contudo, a presença de loops ou inserções dentro da cadeia a proteína exigem uma abordagem diferente das citadas acima e, portanto, existem duas abordagens diferentes para modelagem de loops. A primeira consiste em algoritmos *ab initio*, na qual a conformação do loop é predita por uma busca conformacional guiada por escores e funções de energia. Esse tipo de abordagem leva em conta as posições mais favoráveis termodinamicamente e esse método

pode explorar diferentes representações da proteína, em termos de funções energéticas e algoritmos de otimização e enumeração (PEARSON, 1996). O segundo, utilizado pelo MODELLER, consiste em uma abordagem que utiliza bancos de dados para encontrar um segmento de uma cadeia principal que se encaixe nas extremidades finais dos loops (FISER & SALI, 2003). Essa pesquisa é feita com proteínas de uma maneira geral, não obrigatoriamente homólogas à proteína alvo. Os segmentos que encaixam na região de loop são escolhidos com base em parâmetros geométricos e o melhor é então sobreposto e anelado com a sequência alvo. Por fim, esses segmentos são refinados otimização através de funções energéticas. Esse método foi posto em teste com um número de loops de estruturas conhecidas em ambientes nativos. Loops com 8 resíduos de aminoácidos possuíram 90% de chance de serem modelados com precisão. Já loops com 12 aminoácidos mostraram precisão em 30% dos casos (FISER & SALI, 2003).

Para validar as estruturas modeladas, o programa MODELLER utiliza 3 parâmetros de avaliação: molpdf, DOPE e GA341. O molpdf consiste na somatória de todas as restrições e por isso deve-se selecionar proteínas com o menor desse valor possível. Como o número de restrições de uma proteína depende do seu número de aminoácidos é de se esperar que proteínas maiores tenham maiores valores molpdf. O parâmetro DOPE é um potencial estatístico dependente de distâncias atômicas baseado somente em teorias de probabilidade, onde o modelo mais apropriado será aquele que obtiver menor score. Já o score GA341 utiliza a porcentagem da identidade entre as sequências. O cálculo desse parâmetro se dá através de uma combinação do Z-score calculado com uma função potencial estatística.

Apesar do programa MODELLER já oferecer alguns parâmetros que podem ser utilizados para avaliar as proteínas em questão, outros parâmetros são necessários para confirmação da confiabilidade da estrutura. A ferramenta online *Swiss Model Validation* (<http://swissmodel.expasy.org>) (BIASINI *et al.*, 2014) pode ser utilizada para reforçar a confiabilidade dos modelos moleculares gerados. Essa ferramenta utiliza alguns parâmetros de validação. Um deles é o QMEAN6 (BENKERT *et al.*, 2008), sigla que significa *Qualitative Model Energy ANalysis* (Análise qualitativa da energia do modelo), juntamente com o número 6 pois essa análise qualitativa considerará 6 parâmetros. São eles: energia interação dos carbonos β , somatório energético de todos os átomos, energia de solvatação, energia dos ângulos de torção, concordância com a estrutura secundária e acessibilidade dos solventes. Esse parâmetro não é dependente do tamanho da proteína, pois os parâmetros dependentes do tamanho da proteína são normalizados pelo somatório energético de todos os átomos e pela energia de interação dos carbonos β . Esse parâmetro foi então incorporado a um Z-score que

faz a comparação do modelo a ser avaliado com a média de outras estruturas existentes no banco de dados. O cálculo desse parâmetro é feito pela subtração do QMEAN normalizado e dividindo pelo desvio padrão da distribuição observada. Em uma analogia, cada um dos seis parâmetros do QMEAN também dá origem a um Z-score (BENKERT *et al.*, 2011), sendo todos eles padronizados para que os maiores valores de Z-score indiquem os melhores modelos. Outro parâmetro contido na ferramenta *Swiss Validation* é o PROCHECK (LASKOWSKI *et al.*, 1993), que faz a avaliação estereoquímica da proteína e gera um gráfico detalhado sobre cada resíduo, incluindo a porcentagem deles em cada posição do gráfico de Ramachandran: regiões permitidas, adicionalmente permitidas, generosamente permitidas e proibidas. Outro parâmetro conhecido na avaliação de modelos moleculares é o Verify3D (LUTHY *et al.*, 1992), que testa o quão adequada uma proteína está pela compatibilidade da sua própria sequência com a estrutura, medida por um perfil tridimensional.

Ao avaliar a estrutura modelada da proteína, pode-se prever algumas de suas características e prever o seu comportamento, possibilitando a pesquisa de novas aplicações através das técnicas de *docking* e de dinâmica molecular. Apesar de suas limitações, o *docking* é uma ferramenta poderosa para a descoberta de novos ligantes e revelará se tais moléculas podem se ligar experimentalmente à proteína alvo (CHEN, 2015). Para realizar uma análise aprofundada, simulações de dinâmica molecular podem ser executadas para avaliar a estabilidade das interações, considerando a mudança conformacional do sítio de ligação ao longo da simulação (ADCOCK & MCCAMMON, 2006), o que pode dar ainda mais credibilidade ao dado teórico obtido.

1.5.2 Docking Molecular

Docking ou atracamento molecular é uma técnica que tem como objetivo dar uma predição do complexo ligante-receptor através de métodos computacionais. Os passos gerais para realização desse processo são: amostragem das conformações do ligante no sítio ativo da proteína e avaliação dessas conformações por funções que fornecem um escore. O ideal é que esse primeiro passo gere o máximo possível de conformações do ligante dentro do sítio ativo da proteína e que as funções de escore selecionem a conformação mais semelhante ao que seria observado experimentalmente. Para isso, diversas abordagens são utilizadas para cada um desses quesitos, sempre equilibrando os contrapontos de confiabilidade e custo computacional.

Na bioquímica se conhece duas formas de se representar uma interação. A primeira é o modelo chave-fechadura proposto por Fischer, no qual o ligante se encaixa perfeitamente dentro da proteína, de forma similar ao que acontece quando se insere uma chave dentro de uma

fechadura, sem que ocorram quaisquer mudanças nas conformações. Porém, esse modelo não é tão adequado e hoje se sabe que o modelo que mais se aproxima da realidade é o de ajuste induzido proposto por Koshland, onde a presença de um ligante exerce influência nas cadeias de aminoácidos da proteína e altera, mesmo que de forma sutil, o comportamento da proteína.

Esses modelos são úteis para que se compreenda a forma com que alguns algoritmos calculam a interação proteína-ligante. Quanto menor os graus de liberdade (possibilidade de movimentos das moléculas) dentro do sistema, mais simples será o cálculo do *docking* e consequentemente os resultados serão obtidos de forma mais rápida. Por outro lado, restringir a movimentação do sistema distancia o resultado obtido do que seria observado em condições experimentais. Para fazer o balanceamento entre esses dois contrapontos, existem diferentes algoritmos de *docking* que testam o acoplamento de um ligante rígido em um receptor rígido, um ligante flexível em um receptor rígido, um ligante flexível em um receptor parcialmente flexível (movimentando apenas a parte da molécula envolvida na interação) ou um ligante flexível com um receptor totalmente flexível.

Os algoritmos utilizados para fazer a amostragem dos possíveis encaixes dos ligantes dentro do sítio ativo das proteínas são os seguintes (MENG *et al.*, 2011):

- Algoritmos de Correspondência (BRINT & WILLET, 1987; FISCHER *et al.*, 1993; NOREL *et al.*, 1994):

Se baseia no mapeamento do ligante no sítio ativo considerando formatos e informações químicas. Nesse algoritmo, a proteína e o ligante são representados como farmacóforos, um conjunto de descritores que incluem ligações hidrofóbicas, de hidrogênio e pontos de interações eletrostáticas, definidas por átomos e pontos virtuais (ROY *et al.*, 2015). As distâncias entre os farmacóforos da proteína e do ligante são medidas e calculadas para ligar com átomos correspondentes. A amostragem é então feita através da variação da distância entre os farmacóforos.

- Construção Incremental (RAREY *et al.*, 1996; DESJARLAIS *et al.*, 1986; KUNTZ & LEACH, 1992):

Esse método encaixa o ligante dentro do sítio ativo de forma fragmentada e incremental. Primeiro, o ligante é fragmentado em diversas partes e o fragmento maior, ou de maior importância funcional, é encaixado no sítio ativo, servindo como uma âncora. Em seguida, os outros fragmentos são adicionados de forma incremental. As diferentes amostragens são feitas através de diferentes orientações geradas para cada fragmento, sem mudar a estrutura do ligante em si.

- LUDI (BOHM *et al.*, 1992):

Foca nas interações de hidrogênio e nas interações hidrofóbicas estabelecidas entre a proteína e o ligante. Seu conceito central é constituído por posições espaciais discretas adequadas para formar pontes de hidrogênio ou para preencher sítios hidrofóbicos.

- Monte Carlo (GOODSELL *et al.*, 1993; HART & READ, 1992):

Esse método gera posições do ligante através da flexibilização de ligações, pela translação do corpo rígido ou pela sua rotação. As conformações geradas são testadas por um critério de seleção baseado na energia de interação. Se a conformação passar no critério, ela será salva e modificada para gerar as próximas conformações. As interações continuam a ser testadas até que seja atingido o número de iterações escolhido pelo pesquisador.

- Algoritmos genéticos (MORRIS *et al.*, 1998; JONES *et al.*, 1997; OSHIRO *et al.*, 1995):

Outro método estocástico, esse método tem esse nome pois se baseia na teoria da evolução criada por Darwin. Nesse método, os graus de liberdade do ligante ficam codificados como cadeias binárias chamadas genes. Esses genes compõem um “cromossomo”, que representa a posição do ligante. As variações de conformação acontecem por duas ações chamadas mutações, que acarreta mudanças nos genes, e crossovers, que promovem a troca de genes entre dois cromossomos. Essas ações promovem a amostragem do ligante em diversas conformações que são avaliadas por uma função de escore. No final de uma geração, as conformações que possuem estruturas viáveis sobrevivem e são utilizadas nas próximas gerações.

- Otimização colônia de formigas (KORB *et al.*, 2009) (ACO, do inglês *Ant Colony Optimization*):

Essa metodologia é uma classe de técnicas algorítmicas inspiradas pelo comportamento de formigas fora do formigueiro, onde são capazes de encontrar o caminho mais curto entre seus abrigos e seus alimentos (JIANG & KIM, 1991; CLAUSSEN *et al.*, 2001). Componentes cruciais para essa eficiência das formigas são as trilhas de ferormônios deixadas por elas quando saem do formigueiro e seu comportamento de seguir essa trilha. Sempre que uma formiga precisa escolher entre vários caminhos possíveis, há maior probabilidade de que ela siga pelo caminho com maior concentração de ferormônios.

Os algoritmos ACO utilizam de ferormônios artificiais para simular esse efeito. Esses ferormônios artificiais são representados por valores numéricos que são associados à cada possível componente de solução. Já as “formigas” são processos randomizados que geram uma nova solução candidata a partir de um mecanismo construtivo. O algoritmo PLANTS utilizado nesse trabalho utiliza essa forma de amostragem (KORB *et al.*, 2009).

- Dinâmica molecular (CORNELL *et al.*, 1995; WEINER *et al.*, 1984; BROOKS *et al.*, 1983):

Essa ferramenta pode ser utilizada para se obter resultados de *docking* molecular com representação proteína flexível e ligante flexível, sendo a ferramenta que representa essa relação de forma mais fidedigna quando comparada com outros algoritmos. A dinâmica molecular pode ser muito eficiente quando se quer fazer pequenas mudanças em um ligante para obtenção de um resultado otimizado. Porém essa metodologia não é adequada para se procurar diversos ligantes, pois a mudança energética ocasionada pelo teste de diversas conformações do ligante seria muito grande, o que não é suportado pela dinâmica.

Com objetivo de avaliar as melhores conformações dos ligantes dentro do sítio de ligação da proteína, é necessário que seja atribuído um valor indicador da confiabilidade de cada amostragem. As metodologias utilizadas são as seguintes:

- Funções baseadas em campos de força (KOLLMAN, 1993; AQVIST *et al.*, 2002; CARLSON & JORGENSEN, 1995):

Avaliam a energia de ligação através do somatório de interações hidrofóbicas, descritas pela função potencial de Lennard-Jones, e eletrostáticas, calculadas pelas formulações de Coulomb. Uma desvantagem desse processo é o tempo longo de processamento.

- Funções de escore empíricas (BOHM, 1998; GEHLHAAR *et al.*, 1995; VERKHIVKER *et al.*, 2000; HEAD *et al.*, 1996):

Nessa metodologia, a energia de ligação se decompõe em diversos parâmetros de energia, tais como pontes de hidrogênio, interações iônicas, hidrofóbicas e entropia de ligação. Cada um desses componentes é multiplicado por um coeficiente e o escore é obtido pelo somatório dessa equação. Cada programa possui coeficientes diferentes para fazer o somatório desses parâmetros. O algoritmo PLP utilizado neste trabalho utiliza essa forma de obtenção de score (GEHLHAAR *et al.*, 1995, VERKHIVKER *et al.*, 2000, GEHLHAAR *et al.*, 1995).

- Funções baseadas no conhecimento (ISHCHENKO & SHAKHNOVICH, 2002; FEHER *et al.*, 2003):

Utiliza a análise estatística de estruturas cristalográficas de complexos ligante-proteína para obter as frequências de contatos interatômicos e/ou distâncias entre o ligante e a proteína. Esse tipo de escore está baseado no pressuposto de que quanto mais favorável for uma interação, maior será sua frequência de ocorrência. Uma vantagem dessa abordagem é que ela possibilita a avaliação de interações que são observadas, mas que ainda não são bem compreendidas.

- Escore consenso (CHARIFSON *et al.*, 1999):

Estratégia recente que combina diversos escores diferentes para avaliar os resultados do *docking* molecular. Essa abordagem é interessante pois se algum ligante apresenta escores aceitáveis por diferentes formas de avaliação, é mais provável que ele também se ligue durante o ensaio experimental. Já foi relatado que avaliação da ligação por esse tipo de escore foi capaz de aumentar significativamente o número de acertos em *screens* virtuais (CLARK *et al.*, 2002).

1.5.3 Dinâmica Molecular em GROMACS

A dinâmica molecular é uma técnica que consiste na simulação de átomos esféricos conectados por ligações químicas, onde forças internas experimentadas pelo modelo estrutural estão descritas por funções matemáticas. Utilizando essas leis de comportamento, a dinâmica executa a movimentação do sistema cumprindo com as equações do movimento descritas por Newton, permitindo movimentação estrutural durante o tempo (ADCOCK *et al.*, 2006).

Para que a dinâmica possa ocorrer é necessário que se tenha um campo de força que descreva o comportamento dos átomos e das ligações das moléculas que serão utilizadas na simulação. Os campos de força são construídos a partir de dois componentes, sendo um deles um conjunto de equações (também chamadas funções potenciais) utilizado para gerar as energias potenciais e suas derivadas, as forças (ABRAHAM *et al.*, 2016). O segundo componente é o conjunto de parâmetros utilizados pelas equações. Dependendo do campo de força escolhido, o sistema poderá ser tratado com a representação de átomos unidos (*united-atom*) ou poderá ter todos seus átomos considerados (*all-atom*), bem como terá uma limitação para as moléculas colocadas no sistema. Apenas moléculas cujo comportamento está descrito no campo de força podem ser utilizadas na simulação (ABRAHAM *et al.*, 2016).

Outro componente importante para execução da dinâmica são os arquivos de topologia. Esses arquivos são responsáveis por informar ao GROMACS as combinações de átomos sobre os quais as funções potenciais devem agir. Já as funções potenciais são constituídas pelo potencial de ligação, angular, de torção, Coulombico (eletrostática) e de Lennard-Jones (Van de Waals) (ABRAHAM *et al.*, 2016). Além de se caracterizar as forças do sistema, é necessário que seja aplicado um algoritmo de restrição que delimite os tamanhos das ligações. Para isso, pode-se aplicar o algoritmo *LINear Constraint Solver* (LINCS) (HESS *et al.*, 1997). Para que ligações feitas por distâncias maiores também sejam consideradas, como as eletrostáticas, pode-se aplicar a função *Particle-Mesh Ewald* (PME).

Antes de montar o sistema da dinâmica molecular é necessário que a proteína passe

antes por um processo de minimização energética. Esse processo evita que a proteína inicie a simulação com alta energia de empacotamento, o que geraria uma variação muito brusca no sistema. Essa minimização pode ser feita com dois tipos de algoritmos. O primeiro deles, *steepest descent gradient*, é utilizado para atenuar conflitos altamente não favoráveis. O algoritmo suspende quando o número de determinadas avaliações de energia é executado ou quando os valores absolutos de força (gradiente) chegam à energia determinada. O segundo método para minimização energética é o *conjugate gradient* que, embora seja mais lento do que o *steepest descent* nos primeiros estágios da minimização, se torna mais eficiente próximo ao fim do processo (ABRAHAM *et al.*, 2016).

Após a minimização, os passos de equilibração NVT e NPT são aplicados através do termostato Nose-Hoover para manter a temperatura constante (NOSE, 1984; HOOVER, 1985) e do barostato Parrinello-Rahman para manter o sistema isotrópico a na pressão desejada (PARRINELLO & RAHMAN, 1981). Esses algoritmos são importantes para se considerar um sistema biológico, visto que essas condições se encontram constantes nesses sistemas.

2 OBJETIVOS

2.1 Objetivo geral

O presente trabalho tem como objetivo obter as estruturas primárias e tridimensionais das lectinas de *Arachis duranensis* e *Arachis ipaensis* e analisar interação desses modelos com diferentes carboidratos.

2.2 Objetivos específicos

- Obter as sequências de nucleotídeos que codificam as lectinas de *A. duranensis* e *A. ipaensis*;
- Obter a sequência de aminoácidos das proteínas maduras;
- Obter as estruturas tridimensionais dessas proteínas por modelagem molecular por homologia;
- Obter escores de *docking* molecular dos modelos com carboidratos;
- Avaliar estabilidade e flexibilidade das proteínas na presença ou ausência de ligantes por dinâmica molecular;
- Analisar as interações entre as lectinas e manosídeos ao decorrer do tempo por dinâmica molecular.

3 MATERIAIS E MÉTODOS

3.1 Obtenção da sequência, alinhamento e análise

A ferramenta *tblastn* do NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) foi utilizada tomando a sequência da lectina SL-1 de *Arachis hypogaea* como modelo de busca (Uniprot ID Q70DJ5) para encontrar as sequências de nucleotídeos codificantes das lectinas nos genomas de *Arachis duranensis* (XM_016087184.1) e *Arachis ipaensis* (XM_016319204.1) no banco de dados refseq. Após a obtenção das sequências de nucleotídeos, a ferramenta de tradução do portal de bioinformática ExpASY (<https://web.expasy.org/translate/>) (GASTEIGER *et al.*, 2003) foi utilizada para revelar as sequências de aminoácidos dessa lectinas.

Os peptídeos sinais presentes nessas sequências deduzidas foram detectados e removidos utilizando o servidor SignalP v. 4.1 (<http://www.cbs.dtu.dk/services/SignalP>) (PETERSEN *et al.*, 2011). A similaridade entre as lectinas de *A. duranensis* e *A. ipaensis* foi determinada com a ferramenta Blastp do NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) e essa mesma ferramenta foi utilizada para encontrar a estrutura proteica de maior similaridade com essas lectinas dentro do banco de dados PDB (*Protein Data Bank*) para posteriores comparações. Para realizar o alinhamento das sequências foi utilizada a ferramenta online ClustalOmega (<http://www.ebi.ac.uk/Tools/msa/clustalo>) (SIEVERS *et al.*, 2011) e a análise desse alinhamento foi realizada pela ferramenta ESPript v. 3.1 (<http://espript.ibcp.fr/ESPript/ESPript>). (ROBERT & GOUET, 2014).

As regiões processadas no C-terminal de ambas lectinas foram deduzidas a partir do alinhamento e removidas para execução da modelagem molecular por homologia. O ponto isoelétrico (pI) e a massa molecular teórica das proteínas maduras foram calculados pela ferramenta *Compute pI/MW* do portal Expasy (GASTEIGER *et al.*, 2003) (http://web.expasy.org/compute_pi/) e a busca por sítios de glicosilação foi realizada pelo Servidor NetNGlyc 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>) (JUNG & BRUNAK, 2004).

3.2 Modelagem molecular e validação

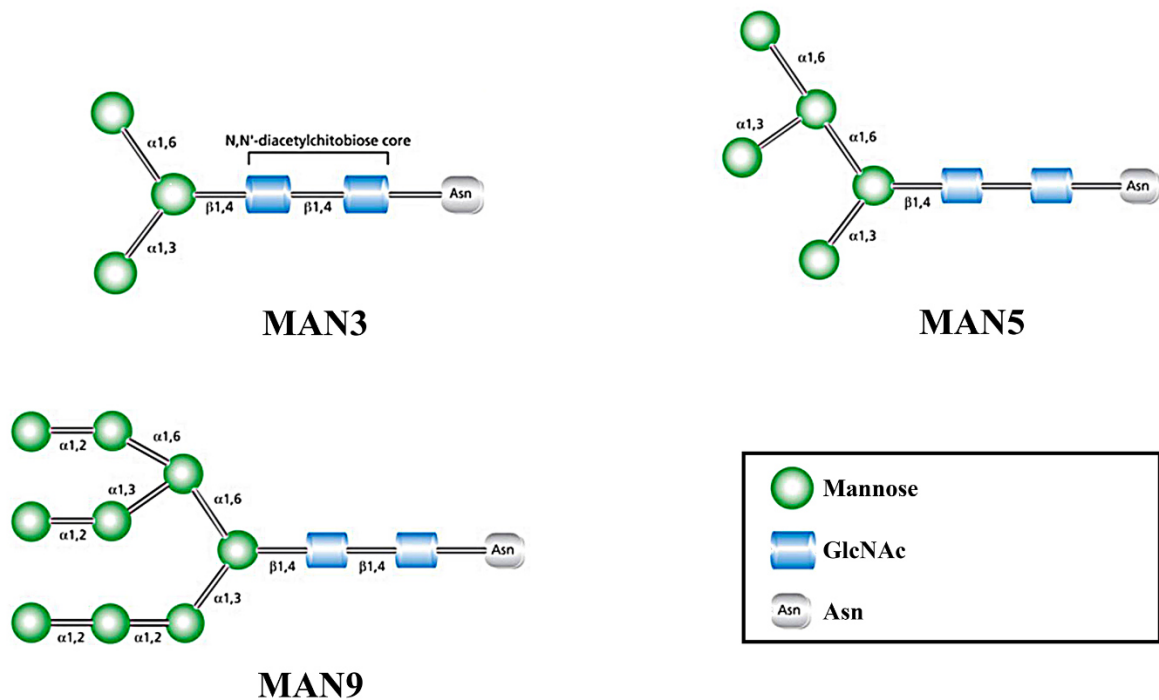
O Software MODELLER v 9.17 foi utilizado para gerar as estruturas tridimensionais das lectinas maduras de *Arachis duranensis* e *Arachis ipaensis*. A ferramenta *protein Blast* do NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) foi utilizada para identificar a

estrutura depositada no PDB (*Protein Data Bank*) de maior homologia com as sequências de lectinas de *Arachis*. A estrutura modelo escolhida foi a lectina de *Centrolobium tomentosum* (CTL, PDB id: 5EYY), uma espécie também pertencente à tribo Dalbergieae. O programa foi configurado para gerar 100 modelos, dos quais 10 foram selecionados pelos melhores escores molpdf, DOPE e GA341 para seguir com outras validações. O melhor modelo foi selecionado com base em parâmetros de avaliação como Qmean6 (BENKERT *et al.*, 2008) e do Z-score (BENKERT *et al.*, 2011), obtidos pela ferramenta online *Swiss Model Validation* (<http://swissmodel.expasy.org>) (BIASINI *et al.*, 2014). As ferramentas online Verify3D (LUTHY *et al.*, 1992) e o PROCHECK (LASKOWSKI *et al.*, 1993) foram utilizadas para avaliar os parâmetros geométricos e a aceitabilidade do ambiente das cadeias laterais de cada aminoácido, disposta no gráfico de Ramachandran. Os modelos que obtiveram melhores parâmetros nas validações foram escolhidos para representar as estruturas das lectinas de *Arachis duranensis* e *Arachis ipaensis*, denominadas ADL e AIL, respectivamente.

3.3 Docking molecular

Os modelos estruturais das lectinas ADL e AIL passaram por uma minimização energética antes de seguir para realização do *docking*. Esse processo foi realizado através do programa UCSF CHIMERA v.1.12 (PETTERSEN *et al.*, 2004), onde a ferramenta *Minimize Structure* foi configurada para execução de 1000 passos de minimização *Steepest Descent* e 100 passos de minimização *Conjugate Gradient*. Em seguida, o *docking* molecular foi realizado com o software CLC Drug Discovery Workbench v. 3.0 (CLC Bio, Boston, MA, USA) considerando as lectinas como receptores rígidos e os ligantes como flexíveis. O atracamento ocorreu no modo *standard precision* (SP), o qual determina as posições favoráveis de ligação e detecta as conformações que se encaixam no sítio de ligação. O algoritmo PLANTSPLP foi utilizado para calcular o escore do *docking* (KORB *et al.*, 2009), no qual os resultados mais negativos indicam uma interação mais favorável. Os sítios de reconhecimento a carboidratos (*carbohydrate binding domain* - CRD) das lectinas foram determinados por sobreposição com a estrutura de CTL (PDB id: 5EYY) e essa coordenada foi utilizada como centro para atracamento dos ligantes. Os parâmetros padrões foram utilizados e o *docking* seguiu com monossacarídeos, dissacarídeos e *N*-glicanos ricos em manose (MAN3, MAN5, MAN9) (Figura 5). O raio de interação foi configurado em 11 Å para monossacarídeos e dissacarídeos e 15 Å para *N*-glicanos ricos em manose. O número de iterações testadas foi de 5000 para monossacarídeos e dissacarídeos e 1000 para os *N*-glicanos ricos em manose.

Figura 5 – Representação estrutural dos N-glicanos ricos em manose utilizados no docking molecular.



3.4 Dinâmica molecular

Para avaliar o comportamento das lectinas ADL e AIL em solução, cada proteína foi submetida a três simulações de dinâmica molecular: uma sem ligante, outra na presença de α -D-manose (MAN) e outra com α -metil-D-manosídeo (MMA). As simulações foram executadas com o pacote de programas GROMACS v. 5.1.2 (ABRAHAM *et al.*, 2015) com os parâmetros do campo de força GROMOS 54a7 (SCHMID *et al.*, 2011). As cargas parciais e a topologia dos ligantes α -D-manose e α -metil-D-manosídeo foram obtidas através do servidor ATB (<https://atb.uq.edu.au/>) (MALDE *et al.*, 2011; KOZIARA *et al.*, 2014) e checadas manualmente. As proteínas foram solvatadas em cubos de 10 Å de aresta com aproximadamente 16000 moléculas de água single point charge (SPC) (BERWEGER *et al.*, 1995) e 9 íons Na^+ para neutralizar os sistemas com ADL e 11 para neutralizar os sistemas da AIL. Em seguida, passos de minimização energética foram realizados com o método *steepest descent* com o critério de convergência de $1 \text{ kJ mol}^{-1}\text{nm}^{-1}$. Após minimização, os passos de equilíbrio NVT e NPT foram aplicados em 100 ps acoplados ao termostato Nose-Hoover para manter a temperatura constante de 300 K (NOSE, 1984; HOOVER, 1985) e ao barostato Parrinello-Rahman para manter o sistema isotrópico a 1 bar (PARRINELLO & RAHMAN, 1981). O algoritmo LINCS (HESS *et al.*, 1997) foi aplicado para restringir todas as ligações antes da simulação, enquanto o método PME (DARDEN *et al.*, 1993) foi utilizado para calcular

interações eletrostáticas de distância máxima de 12 Å e interações de van der Waals de até 15 Å. O software VMD (HUMPHREY *et al.*, 1996) foi utilizado para visualizar a trajetórias e os gráficos foram gerados no software Xmgrace. Utilizou-se também os programas Ligplot (LASKOWSKI & SWINDELLS, 2011) e Pymol (DELANO, 2002) para avaliar as interações ligante-proteína durante cada dinâmica, coletando-se imagens a cada 10 ns de simulação e no momento antes do açúcar sair do sítio.

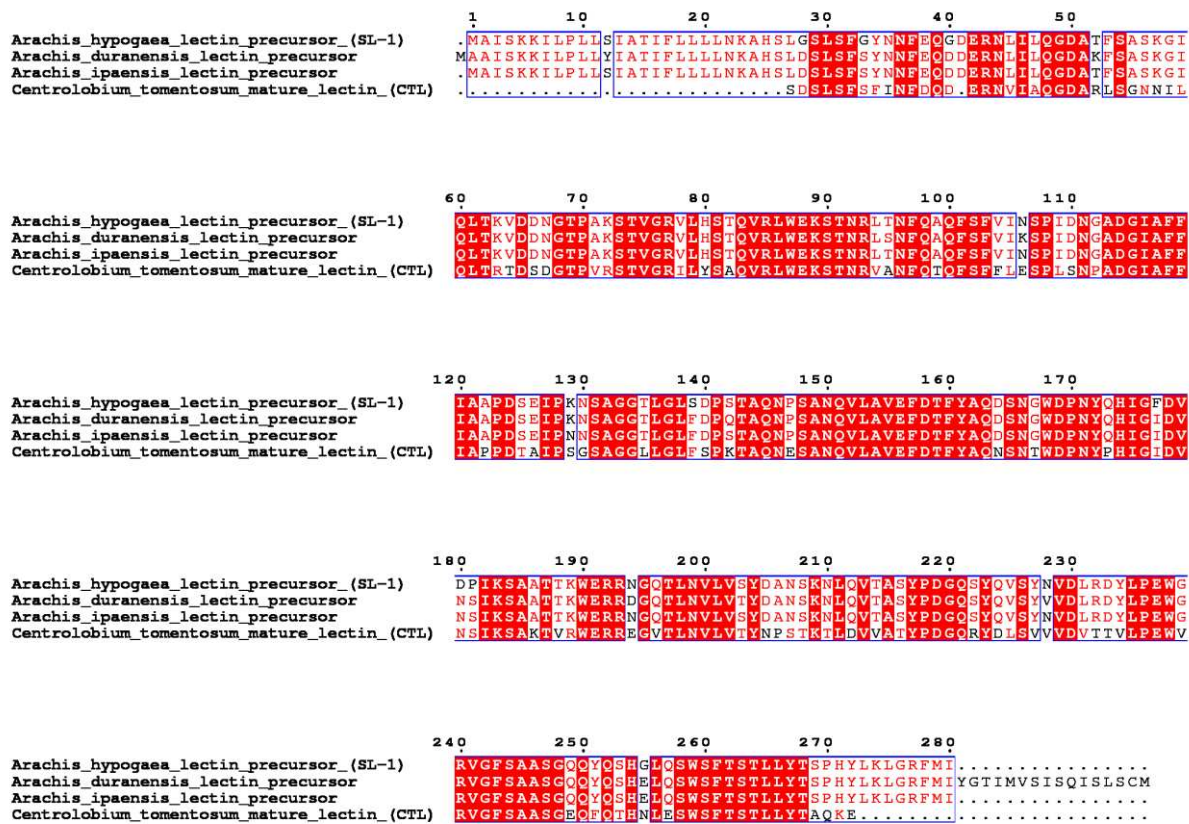
4 RESULTADOS E DISCUSSÃO

4.1 Alinhamento e análise de sequências

Um tblastn foi realizado utilizando a sequência da proteína SL-1 como consulta dentro do banco de dados refseq e assim as sequências de nucleotídeos das lectinas precursoras de *Arachis duranensis* (XM_016087184.1) e de *Arachis ipaensis* (XM_016319204.1) foram adquiridas (BERTIOLI *et al.*, 2016). Um blastp entre as sequências das preproteínas revelou que a pre-ADL e a pre-AIL possuem 95% e 97% de similaridade com a pre-SL-1, respectivamente. O alinhamento das preproteínas de SL-1, ADL e de AIL com a lectina madura de CTL (Uniprot; C0HJX1) (Figura 6), outra lectina da tribo das Dalbergieae, revelou que essas proteínas ainda possuem peptídeo sinal e que elas devem passar por processamentos pós-traducionais para atingirem suas sequências maduras. O servidor SignalP v. 4.1 indicou um peptídeo sinal que se estende do aminoácido 1 ao 27 na pre-ADL e do 1 ao 26 na pre-AIL. Os peptídeos sinais removidos foram N_t-MAAISKKILPLLYIATIFLLLLNKAHS-C_t e N_t-MAISKKILPLLSIATIFLLLLNKAHS-C_t nas respectivas lectinas. Ademais, a comparação entre as sequências das pre-lectinas de *Arachis* com a lectina de CTL revelou também uma sequência não sobreposta na região C-terminal, o que indica um possível processamento pós-traducional nessa região. As sequências removidas nessa região correspondem aos peptídeos N_t-LKLGRFMIYGTIMVSIQSLSLSCM-C_t para pre-ADL e N_t-LKLGRFMI-C_t para pre-AIL, os quais também foram removidos para obtenção das respectivas sequências maduras.

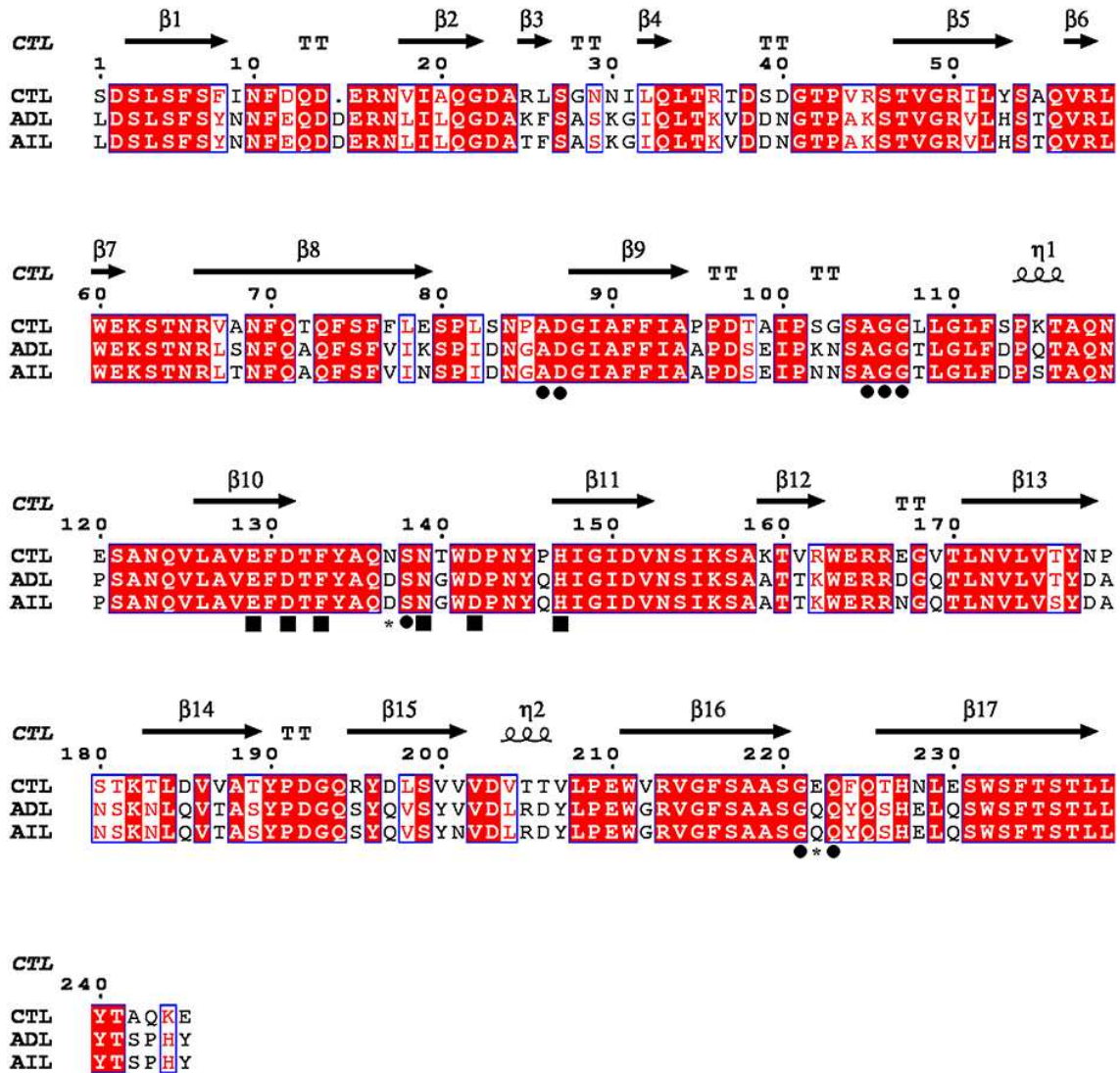
Os tamanhos das proteínas ADL e AIL após clivagem das regiões C e N-terminal estão de acordo com a massa intacta da lectina SL-1 obtida por espectrometria de massas, visto que a sequência alinhada de SL-1 possui massa teórica de 26.636,10 Da correspondente à massa intacta de 26.649,56 Da obtida em estudos prévios (AGRAWAL *et al.*, 2010), onde a pequena diferença entre essas massas pode ser resultante da presença de glicosilações ou isoformas. As proteínas maduras geradas foram então submetidas a um blastp contra proteínas depositadas no PDB e a CTL revelou maior similaridade com a ADL (71%) e com a AIL (70%).

Figura 6 – Alinhamento múltiplo das seqüências de aminoácidos de lectinas precursoras das espécies *Arachis hypogaea*, *Arachis duranensis* and *Arachis ipaensis* com a lectina madura de *Centrolobium tomentosum* (CTL).



As seqüências apresentaram conservação nos resíduos relacionados ao CRD e ao sítio de ligação a metais (*metal binding site*; MBS) presentes na CTL. Exceções foram as substituições dos resíduos Asn137 e Glu222 na CTL por Asp138 e Gln223 em ambas lectinas. Os resíduos conservados em ADL e AIL estão em uma posição a mais devido ao fato de que a CTL possui um aminoácido a menos que essas lectinas na posição 15 do alinhamento (Figura 7). O servidor NetNGlyc 1.0 indicou que as seqüências de aminoácidos de ambas lectinas maduras de *Arachis* possuem sítios de N-glicosilação previstos, porém apresentam uma diferença no sítio de glicosilação, presente como Asn-Glu-Ser em CTL (ALMEIDA *et al.*, 2016) e como Asn-Pro-Ser em AIL e ADL. O servidor NetNGlyc indica uma alerta informando possível impossibilidade de glicosilação nos sítios que contêm prolina após o resíduo de asparagina, o que se explica por um risco de impedimento estérico. Entretanto, dados anteriores (SINGH & DAS, 1994; PATHAK *et al.*, 2006) indicam que a SL-1 contém a mesma modificação e é glicosilada, um forte indício de que ADL e AIL também são.

Figura 7 – Alinhamento múltiplo das seqüências de aminoácidos da lectina de *Centrobium tomentosum* (CTL) com as seqüências maduras das lectinas de *Arachis duranensis* e de *Arachis ipaensis*. Os símbolos correspondem aos seguintes: hélices (η), α -hélice (α), voltas β (TT), resíduos conservados no CRD (\bullet), resíduos envolvidos no sítio de ligação a metal (\blacksquare), resíduos do CRD parcialmente conservados (*).



Os pontos isoeletricos (pI) e as massas moleculares teóricas obtidos para essas proteínas foram 4,72 e 26.690 Da para ADL, e 4,54 e 26.607 Da para AIL, valores próximos àqueles obtidos para CTL, com pI teórico de 4,76 e massa molecular de 27.063 KDa. Dados experimentais da lectina de *Platypodium elegans* (PELa), uma espécie também pertencente à tribo Dalbergieae, mostraram que uma de suas isolectinas possui pI de 5,69 (ARARIPE *et al.*, 2017), um valor próximo ao teórico de 5,73 também indicado pela ferramenta *Compute pI/MW* do portal ExPASy, o que evidencia a confiabilidade desses dados teóricos comparados ao experimental.

4.2 Modelagem molecular e validação

O melhor modelo para cada proteína foi selecionado com base em diversos testes de validação. O melhor modelo de ADL apresentou 92.52% de seus aminoácidos com uma média de escore de correlação 3D-1D maior que 0,2, enquanto o de AIL apresentou 89.76%, o que é um bom parâmetro já que proteínas com mais de 80% são consideradas aceitáveis. A ferramenta de validação Swiss Model indicou um Qmean6 de 0.770 para ADL e de 0.752 para AIL, o que aponta para alta qualidade dos modelos gerados. Os Zscores também foram determinados pela mesma ferramenta e correspondem a 0.013 para ADL e -0.111 para AIL, ambos valores considerados ótimos para modelos moleculares. A análise utilizando PROCHECK mostrou que 100% dos aminoácidos de ambas proteínas estão em regiões permitidas do gráfico de Ramachandran e apontou que nenhum problema sério foi detectado nessas estruturas.

Após validação, as estruturas foram editadas no software Coot (EMSLEY *et al.*, 2010) para remover o peptídeo N_t-TSPHY-C_t que não alinhou com o molde. Embora esse peptídeo tenha sido observado na estrutura primária da lectina de CTL, a mesma não foi observada na estrutura cristalográfica obtida, o que pode ser causado pela limitação da técnica de difração de raios X em captar resíduos com muitos graus de liberdade, como certos loops ou em regiões com glicosilações extensas. As estruturas minimizadas foram sobrepostas (Figura 8) com o modelo de SL-1 obtido do banco de dados ModBase (ID: Q70DJ5) (AGRAWAL *et al.*, 2010) e com CTL (PDB id: 5EYY) para medir o RMSD de seus carbonos alfa em comparação com essas duas outras lectinas, como indicado na Tabela 2.

Tabela 2 – RMSD dos carbonos alfa das estruturas de ADL e AIL superpostas com as estruturas SL-1 e CTL.

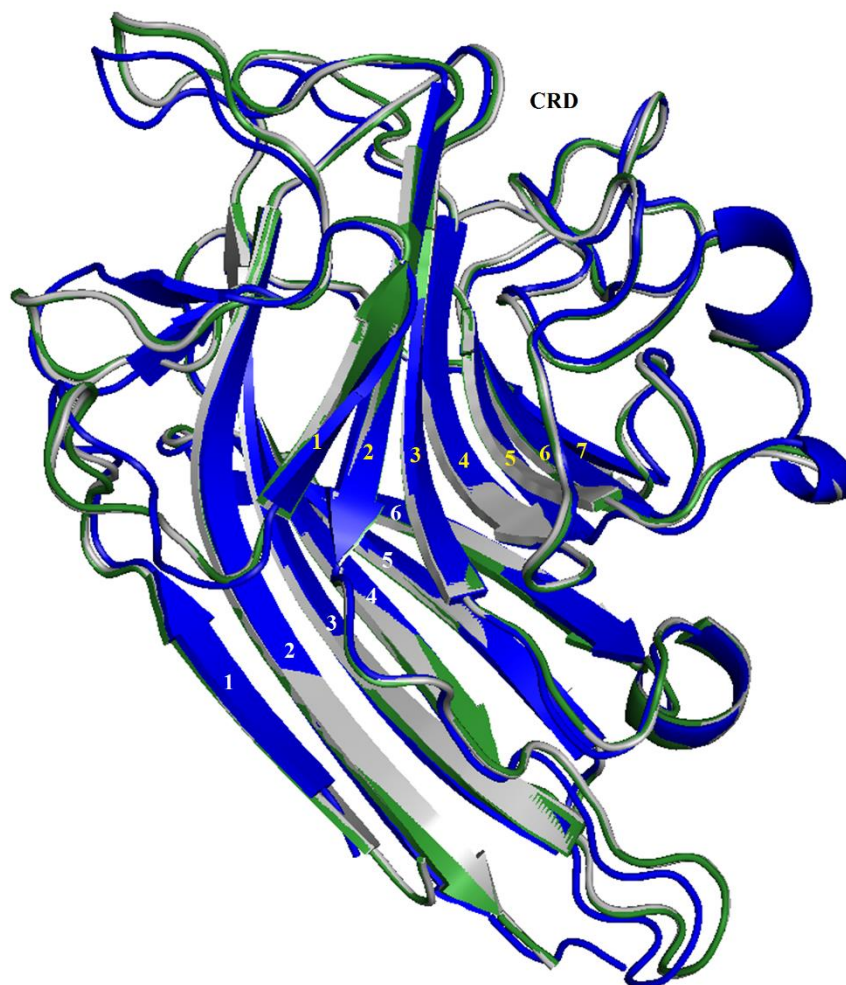
Estruturas sobrepostas	C α RMSD
AIL - SL-1	0.504
ADL - SL-1	0.532
AIL - CTL	0.373
ADL - CTL	0.385

Fonte: Elaborada pelo Autor

O modelo molecular da lectina SL-1 utilizado para fazer essa sobreposição foi obtido através da modelagem comparativa da sua sequência utilizando a estrutura da lectina de *Pterocarpus angolensis* (PATHAK *et al.*, 2006), a qual possui 65% de similaridade com SL-1. O fato de que os modelos de ADL e de AIL foram gerados utilizando uma estrutura de maior

homologia, reforça a maior fidelidade desses modelos às estruturas reais, o que explica o RMSD discrepante dessas estruturas quando comparadas com o modelo de SL-1. Os modelos gerados possuem o mesmo motivo *jellyroll* observado em muitas proteínas de capas virais (ARGOS *et al.*, 1980; CHELVANAYAGAM *et al.*, 1992) e em muitas lectinas de leguminosas, descrito pela sua composição de duas folhas-beta anti-paralelas (LORIS *et al.*, 1998). Em ambas lectinas, uma folha-beta curvada de 7 fitas se posiciona voltada para o CRD (Figura 8) mantendo os loops 86-88, 105-109, 134-140 e 221-224, os quais contêm aminoácidos diretamente envolvidos com o sítio de ligação (Figura 8). Do outro lado, uma folha-beta de 6 fitas parcialmente estendida mantém a estabilidade estrutural da proteína.

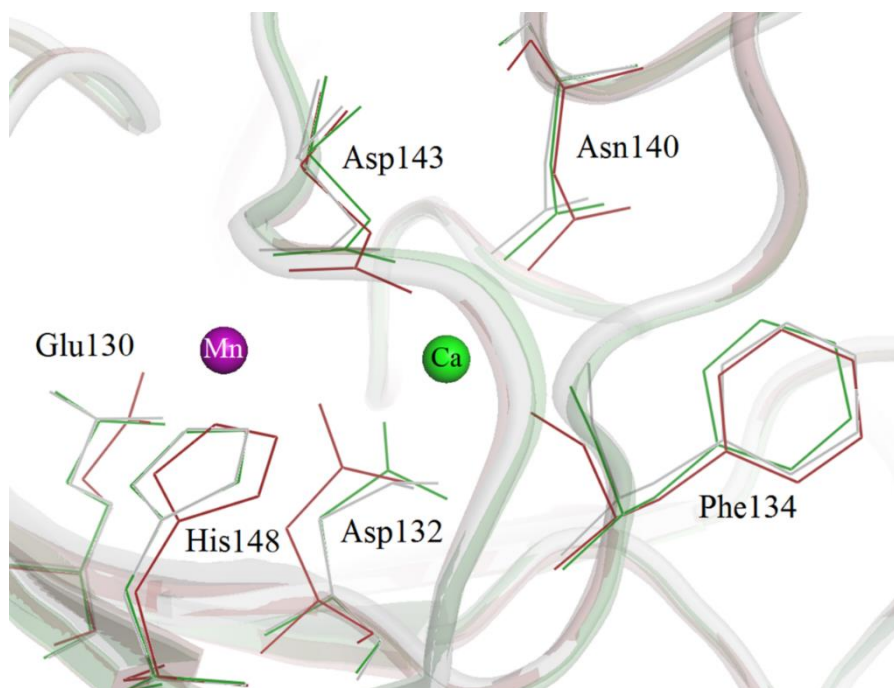
Figura 8 – Sobreposição das estruturas de ADL (verde) e de AIL (cinza) com SL-1 (azul) na representação tipo cartoon. Números brancos enumeram as fitas da folha β traseira enquanto os amarelos enumeram as fitas da folha β frontal, voltada para o CRD.



Os loops envolvidos no CRD são similares na ADL e na AIL, mas alguns se diferem da CTL, como pode ser observado no alinhamento da Figura 6. Os aminoácidos presentes nos loops 128-135, 139-143 e 155-157 fazem contato direto com os íons de cálcio e manganês na estrutura da CTL (PDB id 5EYY) e esses resíduos estão todos conservados nas lectinas de *Arachis* (Figura 9. Dentre os resíduos presentes nos loops que formam o arcabouço do MBS, os aminoácidos Glu130 e His148 interagem diretamente com o Mn^{2+} e os aminoácidos Asp132, Phe 134, Asn140 e Asp143 interagem diretamente com o Ca^{2+} .

Assim como outras lectinas de leguminosas, ADL e AIL possuem um aminoácido ligado por uma ligação peptídica *cis*, localizada entre Ala87 e Asp88. Essa conformação é mantida através de uma molécula de água estrutural que conecta um dos metais divalentes (Ca^{2+}) ao resíduo de ácido aspártico, estabilizando assim a conformação *cis*. Embora essa característica seja bem conhecida em lectinas de leguminosas pela cristalografia de raios X, a falta de campos de força que caracterizem esses íons ainda é um fator limitante para o estudo dessas interações pelo pacote GROMACS. O resíduo correspondente a Asp88 nas lectinas de *Arachis* é altamente conservado em outras lectinas de leguminosas e ele é muito importante para estabilizar o carboidrato no CRD, estando presente na base do sítio de ligação (LORIS *et al.*, 1998).

Figura 9 – Representação do sítio de ligação a metal conservado da CTL (vermelho) superposto com estruturas modeladas ADL (verde) e AIL (cinza). Tanto o Ca^{2+} quanto o Mn^{2+} foram observados a partir da estrutura cristalográfica de CTL (PDB id: 5EYY).



4.3 Docking molecular

Os escores de *docking* obtidos para ADL e AIL com monossacarídeos, dissacarídeos e N-glicanos ricos em manose estão dispostos na Tabela 3. AIL apresentou melhor escore nas suas interações com monossacarídeos e dissacarídeos, com exceção dos dimanosídeos formados por ligações $\alpha 1 \rightarrow 3$, $\beta 1 \rightarrow 3$ e $\beta 1 \rightarrow 6$. Em relação aos monossacarídeos, a presença do grupo metil no ligante α -metil-D-manosídeo aumentou o valor de escore de atracamento com as duas lectinas, corroborando com dados experimentais previamente obtidos para SL-1 (AGRAWAL *et al.*, 2010). O tipo de ligação glicosídica dos dimanosídeos causou diferenças nos valores dos escores, seja aumentando ou diminuindo, porém, a presença ou ausência do grupo metil não gerou diferenças significativas nesses valores. A ADL manifestou escores de atracamento melhores com os N-glicanos MAN3 e MAN9, enquanto os escores contra MAN5 foram relativamente baixos para as duas proteínas. Esses glicanos ricos em manose são relevantes em alguns processos moleculares como os de reconhecimento celular, recrutamento de leucócitos (GU *et al.*, 2012), auxílio no *folding* de proteínas quando interagem com lectinas (SCHWARZ *et al.*, 2011). MAN3 constitui o centro de todos os N-glicanos e MAN5 é o precursor de diversos glicanos ricos em manose, enquanto MAN9 pode ser encontrado em proteínas como a gp120 do envelope viral do HIV ou como receptores de insulina (STANLEY *et al.*, 2009). Esse comportamento foi previamente observado para outras lectinas de leguminosas específicas a manose como as das espécies: *Pisum arvense*, *Pisum sativum*, *Lens culinaris*, *Vicia faba*, *Dioclea lasiophylla* e *Dioclea reflexa* que apresentam efeitos vasorelaxantes, proinflamatório e anti-HIV (STANLEY *et al.*, 2009; HANSEN *et al.*, 1989; PINTO-JUNIOR *et al.*, 2017a, 2017b, 2017c), o que sugere que ambas ADL e AIL podem apresentar essas mesmas atividades biológicas.

Tabela 3– Escore de atracamento da ADL e da AIL com os carboidratos testados no screen virtual. a) Escore do docking = $E_{inter} + E_{intra}$, onde E_{inter} é a energia de interação ligante-proteína.

$$E_{inter} = \sum_{i \in ligand} \sum_{j \in ligand} [E_{PLP}(r_{ij}) + 332.0(q_i q_j / 4r_{ij}^2)]$$

Carboidratos	Escore do Docking ^{a)}	
	ADL	AIL
α -metil-D-manosídeo	-38.24	-40,45
2-(Acetil amino)-2-deoxi- α -D-manopiranosose	-37,24	-38,48
N-acetil-D-glicosamina	-37,88	-39,26
β -D-manose	-36,85	-38,40
α -D-manose	-36,75	-40,45
Manose-1,2- α -D-manose	-47,94	-48,99

Manose-1,3- α -D-manose	-49,37	-46,66
Manose-1,4- α -D-manose	-45,41	-46,12
Manose-1,6- α -D-manose	-48,94	-51,99
Manose-1,2- β -D-manose	-46,20	-46,43
Manose-1,3- β -D-manose	-49,02	-46,85
Manose-1,4- β -D-manose	-42,73	-46,64
Manose-1,6- β -D-manose	-49,49	-47,73
Metil manose-1,2- α -D-manose	-50,51	-48,71
Metil manose-1,3- α -D-manose	-52,26	-49,57
Metil manose-1,4- α -D-manose	-46,77	-46,15
Metil manose-1,6- α -D-manose	-46,98	-51,42
Metil manose-1,2- β -D-manose	-47,19	-44,94
Metil manose-1,3- β -D-manose	-49,40	-48,02
Metil manose-1,4- β -D-manose	-41,27	-44,57
Metil manose-1,6- β -D-manose	-48,59	-47,81
MAN3	-56,06	-46,80
MAN5	-37,67	-38,96
MAN9	-63,15	-48,96

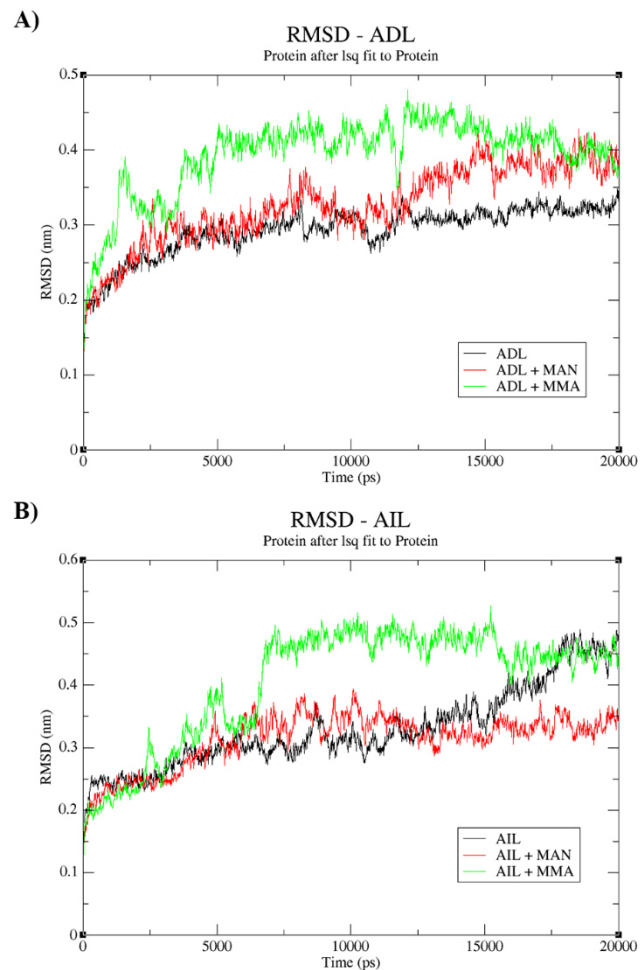
Fonte: elaborada pelo autor.

4.4 Dinâmica molecular

4.4.1 Desvio quadrático médio e flutuação quadrática média

Gráficos do desvio quadrático médio (RMSD) foram plotados para avaliar a estabilidade da proteína durante todas as simulações: sem ligante, com D-manose e com α -metil-D-manosídeo. Para ADL, a variação atingiu equilíbrio em volta de 4.000 ps (Figura 10 A). É importante considerar esse intervalo pois ele representa o tempo que a proteína leva para fazer sua transição do estágio minimizado para a sua conformação totalmente solvatada. O gráfico gerado para proteína sem ligante sofreu menos variações comparado com as simulações na presença de ligantes. D-manose abandonou o sítio de ligação da proteína a proteína após 9.500 ps, o que poderia explicar o aumento ocorrido no RMSD após 10.000 ps durante essa simulação. O α -metil-D-manosídeo, por outro lado, permaneceu interagindo com a proteína durante toda simulação.

Figura 10– Gráficos do desvio quadrático médio da proteína durante as simulações de dinâmica molecular. A) ADL; B) AIL.

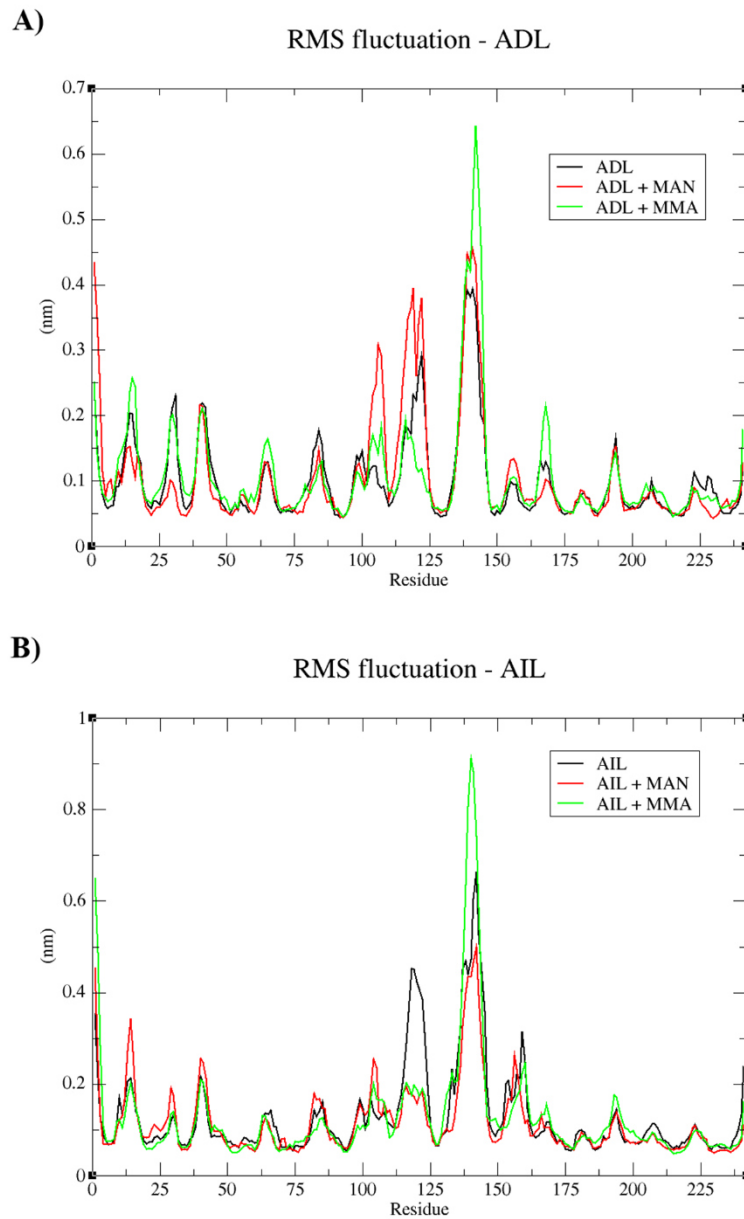


Fonte: Elaborada pelo autor.

AIL levou mais tempo para chegar ao equilíbrio e seu RMSD estabilizou por volta de 7.000 ps (Figura 10 B). A simulação feita sem ligante manifestou uma elevação no desvio após 15.000 ps, mas nenhuma diferença foi vista no sistema. Simulações revelaram que a D-manose deixou de interagir com a proteína após 2,720 ps e que α -metil-D-manosídeo se soltou após 15.170 ps.

Os fatores-B teóricos desses sistemas foram gerados para avaliar a flexibilidade de cada resíduo de aminoácido durante as simulações. Resultados dos gráficos da flutuação quadrática média (RMSF; do inglês *Root Mean Square Fluctuation*) (Figura 11) revelaram que os aminoácidos presentes nas regiões 132-142 foram os que mais variaram posição em ambas lectinas. A região 114-123 flutuou na ausência de ligante e foi estabilizada pelas interações carboidrato-lectina. Na ADL (Figura 11 A), a D-manose não pôde estabilizar a região 114-123 e a região 103-108 se comportou de forma diferente quando comparada às outras simulações.

Figura 11– Gráficos da flutuação quadrática média de cada resíduo de aminoácido durante as simulações de dinâmica molecular. A) ADL; B) AIL

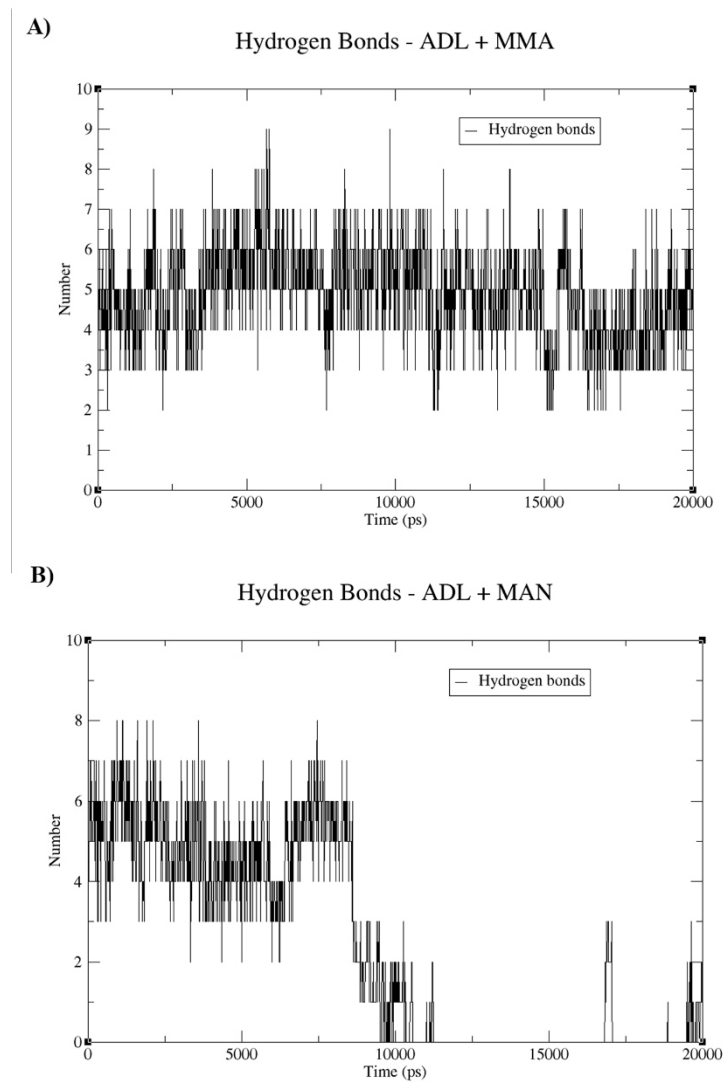


Fonte: Elaborada pelo autor.

A análise das pontes de hidrogênio revelou que a ADL foi capaz de estabelecer mais pontes de hidrogênio com MMA (Figura 12 A) do que com MAN (Figura 12 B). Como discutido anteriormente, MAN se separou do sítio de ligação da ADL após 9.500 ps e as pontes de hidrogênio estabelecidas após esse momento aconteceram com aminoácidos laterais carregados que não pertencem ao CRD. Nas simulações da AIL, MAN se separou do sítio de ligação após 2.700 ps e as interações de hidrogênio estabelecidas após esse momento também

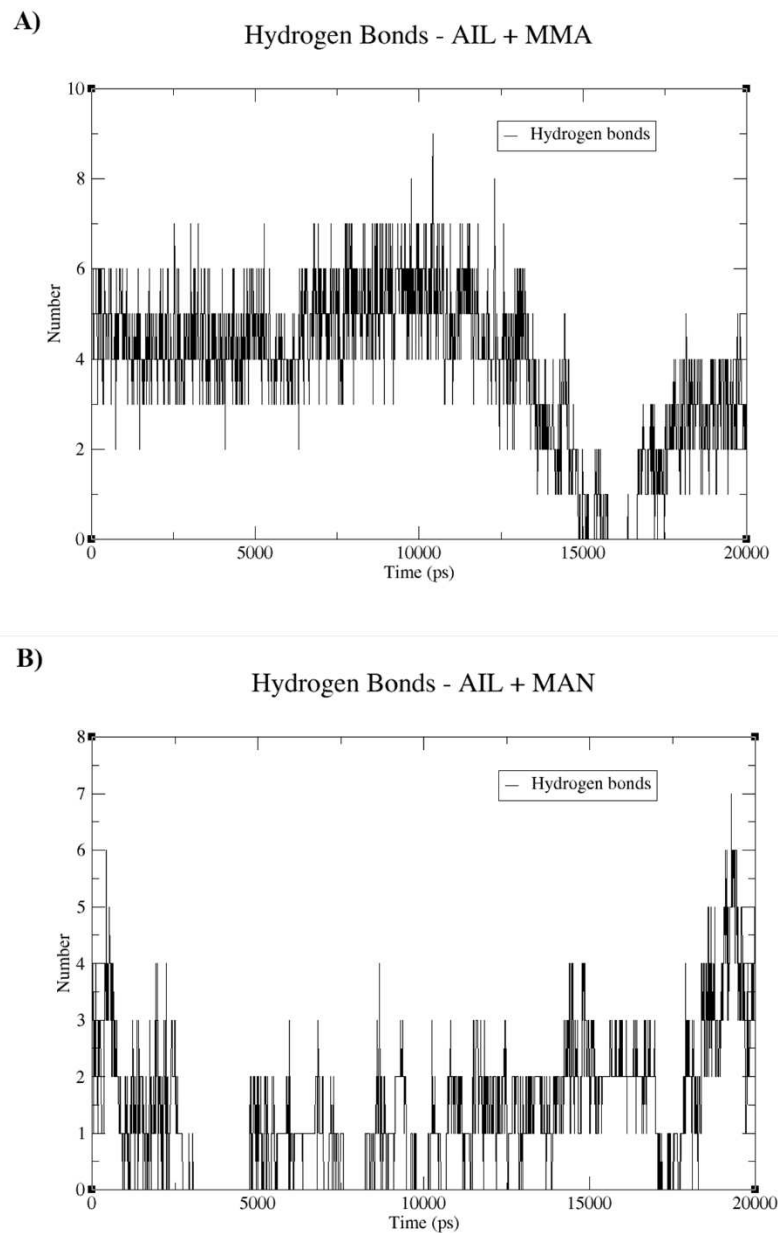
ocorreram de forma inespecífica com aminoácidos polares, o que indica a menor especificidade da AIL com MAN (Figura 13 B). Na simulação AIL-MMA, também houve escape do ligante, mas somente após 15.170 ps, um indício da importância do grupo metil para estabilizar interação (Figura 13 A).

Figura 12– Número de ligações de hidrogênio estabelecidas entre ADL e seus ligantes em cada momento da dinâmica molecular. A) MMA; B) MAN



Fonte: Elaborada pelo autor.

Figura 13 – Número de ligações de hidrogênio estabelecidas entre AIL e seus ligantes em cada momento da dinâmica molecular. A) MMA; B) MAN



Fonte: Elaborada pelo autor

4.4.2 Análise de interações

As interações estabelecidas entre as lectinas e seus respectivos ligantes durante as simulações foram avaliadas através de representações coletadas em intervalos de 10 ns ou no momento anterior à separação ligante-proteína. As representações foram feitas utilizando o programa PyMol para visualização em 3D da interação e com o programa LigPlot para visualização em 2D, incluindo a influência dos resíduos hidrofóbicos. Os dados coletados da

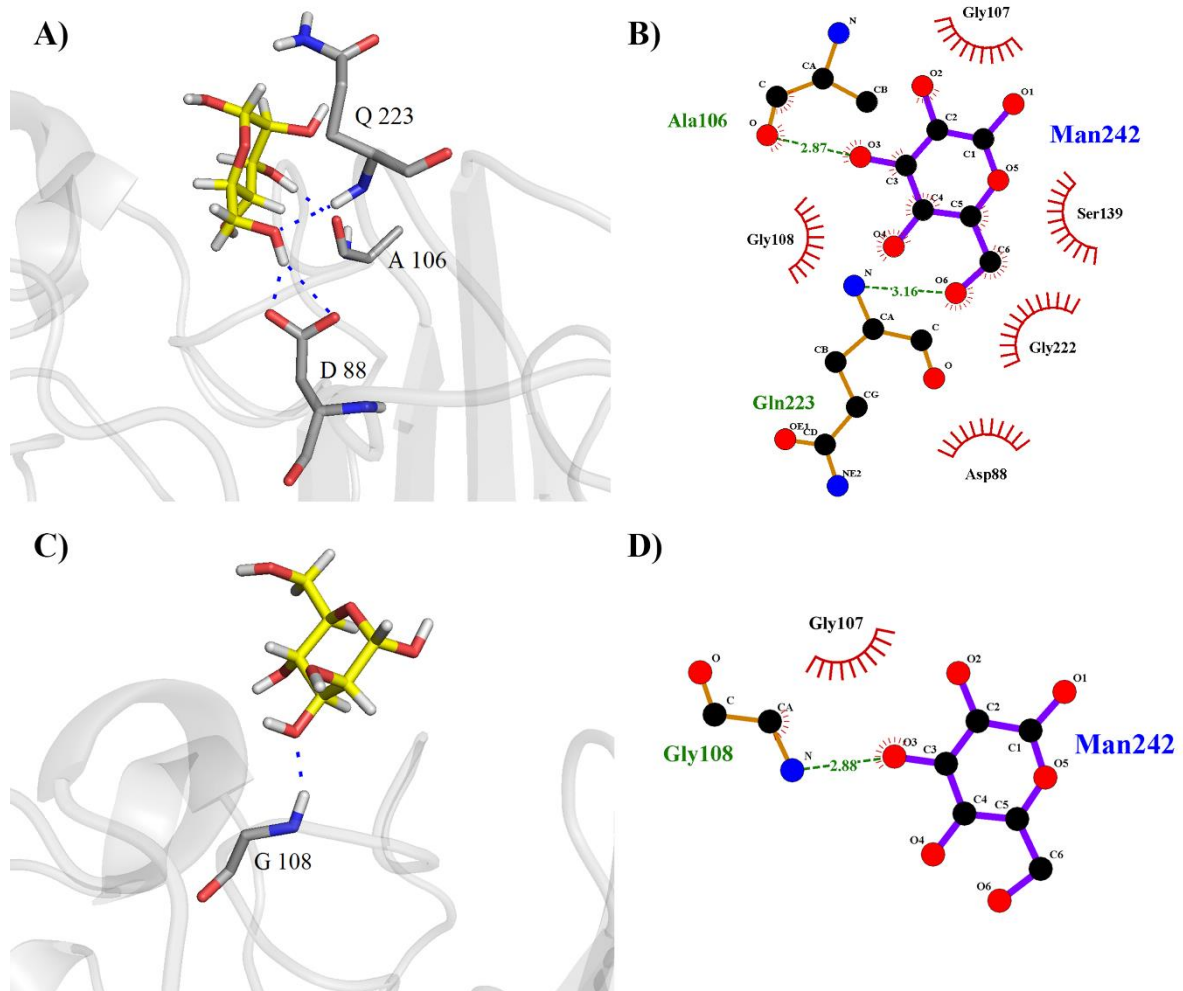
ADL e da AIL revelaram a conservação dos aminoácidos que formaram ligações de hidrogênio com os carboidratos testados.

No primeiro momento, 0 ns, de interação entre AIL e MAN (Figura 14 A-B), pode-se observar que as ligações de hidrogênio são estabelecidas somente com os aminoácidos Asp88, Ala106 e Gln223, enquanto as interações hidrofóbicas ocorrem com Gly107, Gly108, Ser139 e Gly222. Quando a MAN está prestes a se separar, a 2.72 ns (Figura 14 C-D), os únicos aminoácidos interagindo são Gly107 e Gly108, ambos posicionados na lateral do sítio. O fato dessas ligações não terem segurado o ligante no sítio de ligação pode ser atribuído pela distância entre o aminoácido basal Asp88 e o ligante.

Para MMA, o ligante permaneceu em contato com a lectina por 15.17 ns. As imagens coletadas revelaram que o MMA estabeleceu mais ligações de hidrogênio com os aminoácidos no momento 0 ns (Figura 15 A-B), incluindo Asp88, Ala106, Gly108, Gln223, Gln224, e interações hidrofóbicas com a participação dos resíduos Ala87, Gly107, Phe134, Ser139 e Gly222. Em 10 ns (Figura 15 C-D), uma flexibilização dos loops do CRD é observada, o que resultou na formação de novas pontes de hidrogênio com Ala87 e com Gly222, ao passo que a interação envolvendo o resíduo Gly108 foi desfeita. Gly107, Gln222 e Gln223 também atuam sobre o ligante por interações hidrofóbicas. Para MMA, as imagens coletadas em 0 e 10 ns não revelaram interações hidrofóbicas com o grupamento metil (Figura 15 B e D; espinhos em vermelho) e uma diferença pode ser observada na orientação do ligante quando ele está se separando do CRD.

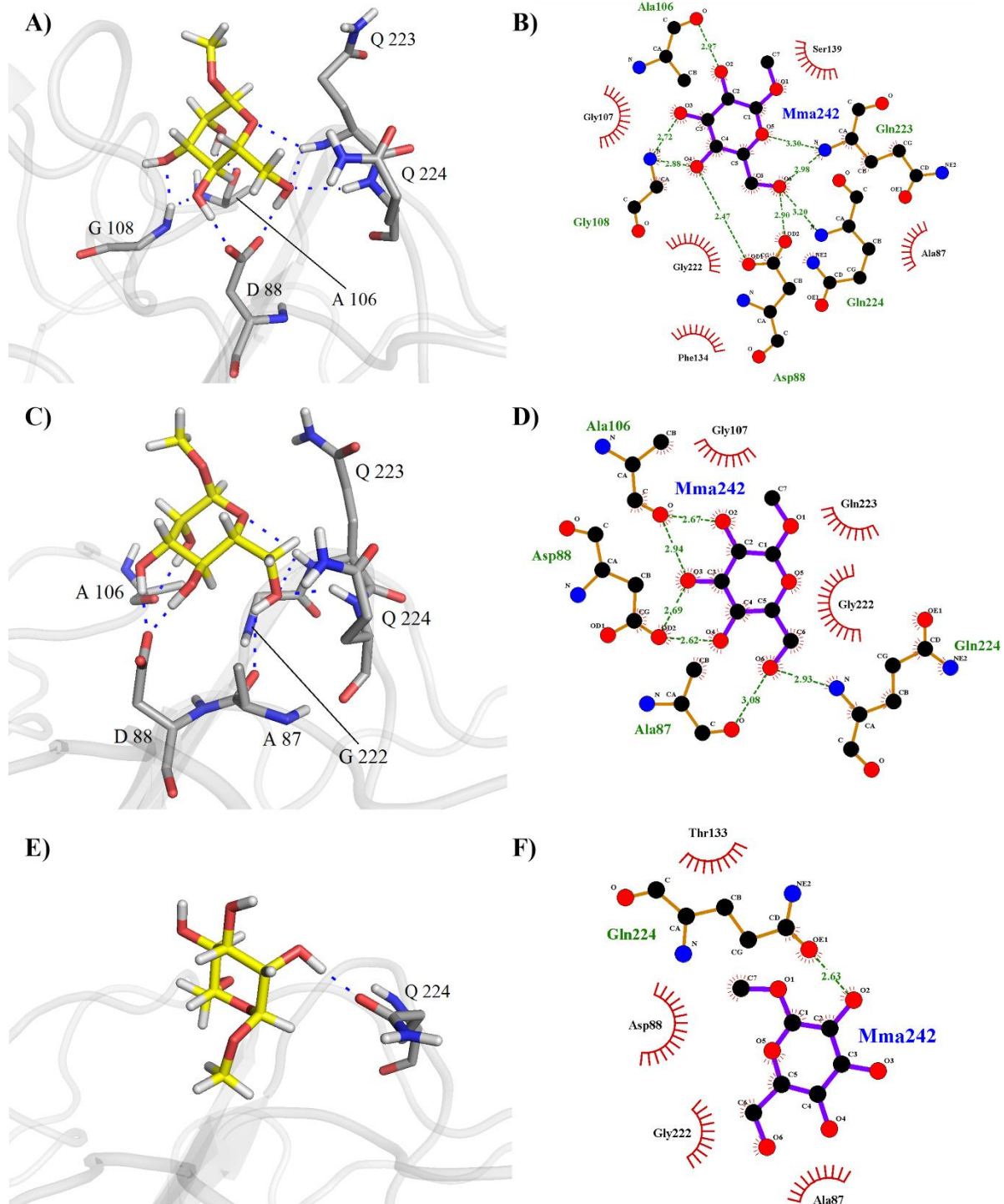
No caso da MAN, o carbono 1 (C1) tende a ficar virado para o lado contrário do CRD, para que assim o C6 e o oxigênio 6 (O6) possam interagir com os aminoácidos basais do CRD. Na simulação com MMA, o ligante permanece na mesma posição, a qual pode ser reforçada pelo grupo metil que estabelece interações com as cadeias carbônicas dos resíduos laterais Gln223, Gln224, Tyr133, Gly86 e Asn85. Ademais, a presença desse grupamento metil, ausente na D-manose, diminui a liberdade de movimentação do ligante e assim dificulta sua rotação, mantendo o contato com os aminoácidos basais e, conseqüentemente, mantendo o C1 e o C7 (do grupo metil) afastados da base do sítio. O C1 do MMA é observado em uma orientação diferente somente quando está se separando do CRD, como pode se observar no momento 15.17 ns (Figura 15 E-F), quando o grupo metil está estabelecendo interações hidrofóbicas com Ala87, Asp88, Thr 133 e Gly222, mas somente uma ligação de hidrogênio com Gln224 (Figura 15 F).

Figura 14 – Domínio de reconhecimento a carboidratos da AIL interagindo com MAN durante a simulação de dinâmica molecular. Esquerda: representações geradas pelo programa PyMol. Tracejados azuis representam pontes de hidrogênio. Direita: representações geradas pelo programa LIGPLOT com pontes de hidrogênio representadas em verde e interações hidrofóbicas representadas em vermelho. A e B) 0 ns; C e D) 2.72 ns.



Fonte: Elaborada pelo autor.

Figura 15 – Domínio de reconhecimento a carboidratos da AIL interagindo com MMA durante a simulação de dinâmica molecular. Esquerda: representações geradas pelo programa PyMol. Tracejados azuis representam pontes de hidrogênio. Direita: representações geradas pelo programa LIGPLOT com pontes de hidrogênio representadas em verde e interações hidrofóbicas representadas em vermelho. A e B) 0 ns; C e D) 10 ns; E e F) 15.17 ns.



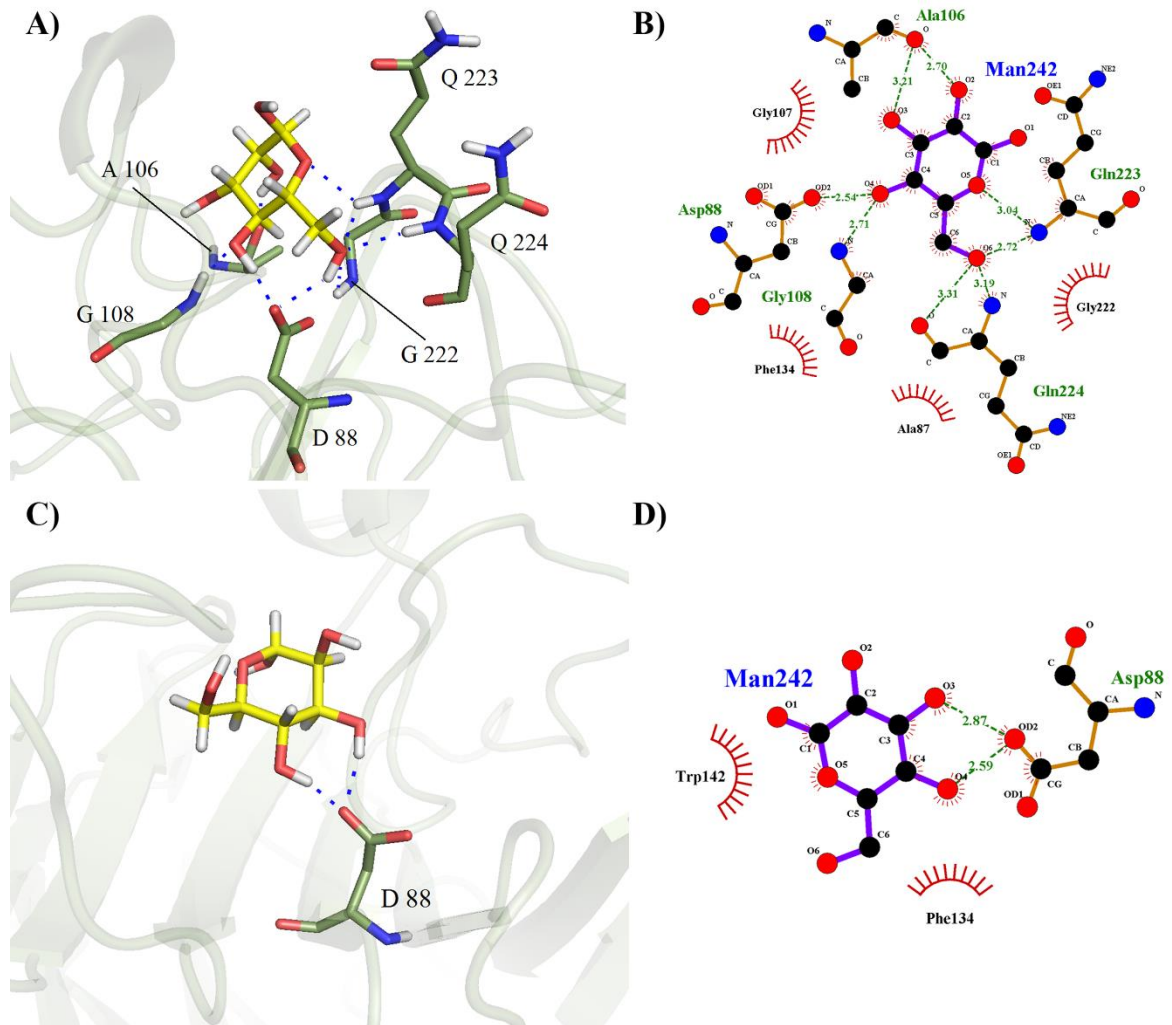
Fonte: Elaborada pelo autor.

Uma orientação similar no ligante é observada quando MAN interage com ADL. Embora mais ligações de hidrogênio sejam estabelecidas, C1 e O1 permanecem orientados para o lado de fora do sítio de ligação. C6 e O6, por outro lado, se posicionam próximos aos resíduos Asp88, Gln223 e Gln224. Em 0 ns (Figura 16 A-B), a lectina apresenta interação com os resíduos Asp88, Ala106, Gly108, Gly222, Gln223 e Gln224 por pontes de hidrogênio e com Ala87, Gly107, Phe134 e Gly222 por interações hidrofóbicas. Ainda assim, a ligação não demonstrou estabilidade já que o ligante se separou após 9.5 ns (Figura 16 C-D). No momento da separação, MAN está interagindo com Asp88 por ponte de hidrogênio e com Phe134 e Trp142 por interações hidrofóbicas.

MMA apresentou maior estabilidade (Figura 17) e esse ligante permaneceu ligado por toda simulação. Ao contrário da AIL, o grupo metil (C7) do MMA conseguiu estabelecer interações hidrofóbicas com alguns resíduos, incluindo Asp138, Ser139, Gln223 e Gln224. Essas interações revelaram ser importantes pois o C7 não alterou abruptamente sua orientação como foi observado para AIL e, assim, manteve-se ligado durante toda a simulação. Durante a simulação, os aminoácidos Ala87, Asp88, Ala106, Gly108, Gly222, Gln223 e Gln224 fizeram interações de hidrogênio com MMA e os resíduos Ala87, Ala106, Gly107, Gly108, Phe134, Asp138, Ser139, Gly222 e Gln223 estabeleceram interações hidrofóbicas. A Tabela 4 apresenta um sumário dos aminoácidos que interagiram com os ligantes estudados durante as simulações.

Figura 16 – Domínio de reconhecimento a carboidratos da ADL interagindo com MAN durante a simulação de dinâmica molecular. Esquerda: representações geradas pelo programa PyMol. Tracejados azuis representam pontes de hidrogênio. Direita: representações geradas pelo

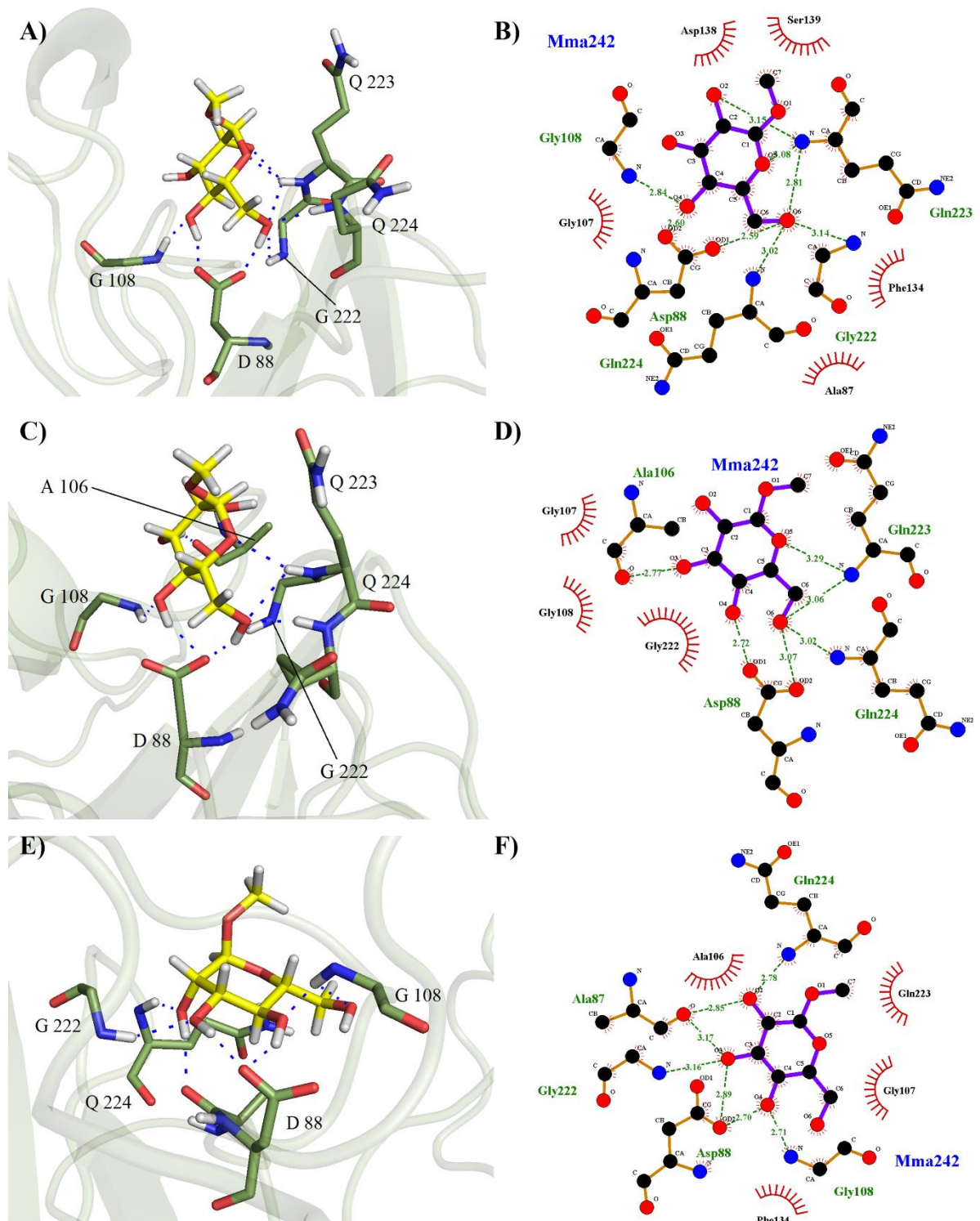
programa LIGPLOT com pontes de hidrogênio representadas em verde e interações hidrofóbicas representadas em vermelho. A e B) 0 ns; C e D) 9.5 ns.



Fonte: Elaborada pelo autor.

Figura 17 – Domínio de reconhecimento a carboidratos da ADL interagindo com MMA durante a simulação de dinâmica molecular. Esquerda: representações geradas pelo programa PyMol. Tracejados azuis representam pontes de hidrogênio. Direita: representações geradas pelo

programa LIGPLOT com pontes de hidrogênio representadas em verde e interações hidrofóbicas representadas em vermelho. A e B) 0 ns; C e D) 10 ns; E e F) 20 ns.

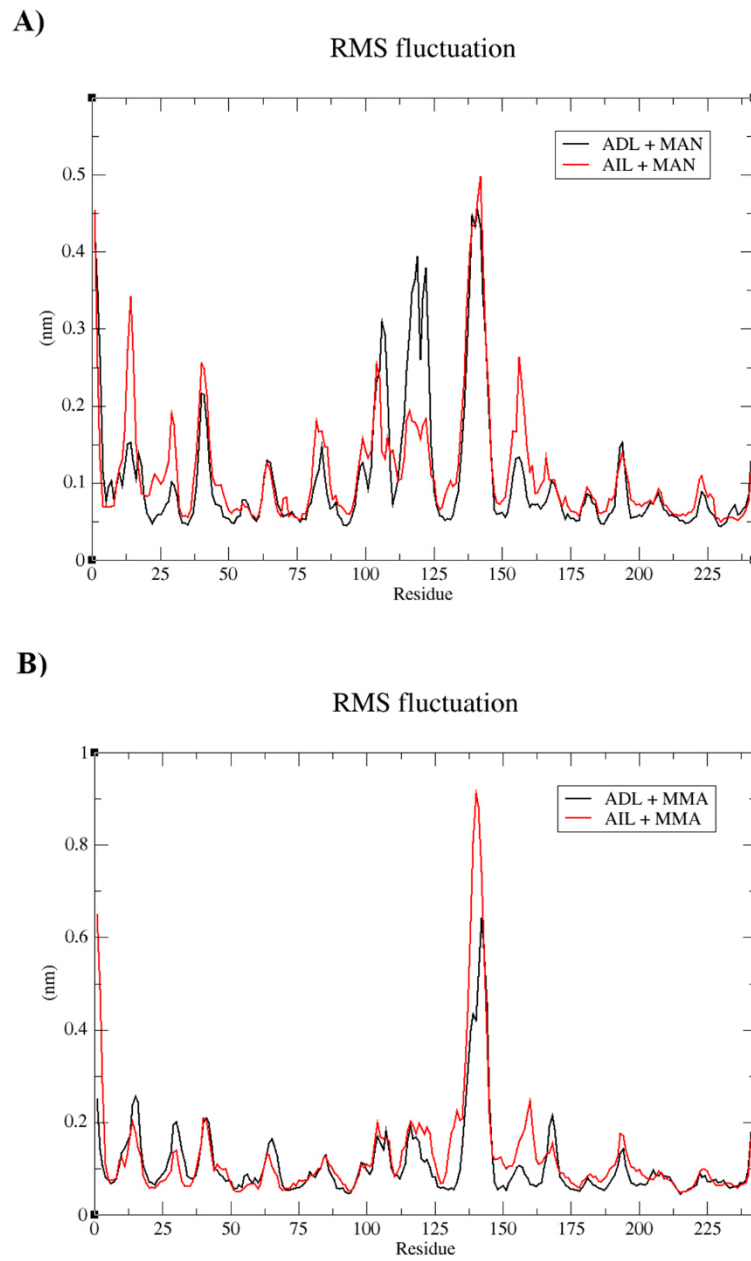


Fonte: elaborada pelo autor.

Como observado na Figura 16, os aminoácidos Asn103 e Ser116 da AIL, os quais estão localizados em loops próximos ao sítio de ligação, estão substituídos por Lys103 e Gln116 na ADL, os quais possuem grupos-R maiores. Essa mudança pode ter causado algumas das

diferenças observadas nas interações proteína-ligante, através da mudança de mobilidade dos resíduos Ala106 e Gly108, os quais foram observados formando pontes de hidrogênio com ambos MAN e MMA. Para ilustrar essa hipótese, um gráfico de RMSF foi gerado (Figura 18) comparando a ADL e a AIL na presença dos ligantes. Os resultados mostraram que a região correspondente a Asn103-Ser116 flutuou mais em ADL que em AIL, indicando que os grupos-R mais volumosos dos aminoácidos Lys103 e Gln116 podem ter influenciado a mobilidade do loop assim como a estabilidade das interações com os ligantes.

Figura 18 – Comparação dos gráficos da flutuação quadrática média de cada aminoácido das lectinas ADL e AIL. A) Complexadas com MAN; B) Complexadas com MMA



Fonte: elaborada pelo autor

Tabela 4– Resíduos de aminoácidos das proteínas ADL e AIL que interagem com os ligantes MAN e MMA durante as simulações de dinâmica molecular.

Tipo de interação	Ligante	ADL	AIL
Interações Hidrofóbicas	MAN	Ala87, Gly107, Phe134, Trp142, Gly222,	Ala87, Gly107, Phe134, Trp142, Gly222
	MMA	Ala87, Ala106, Gly107, Gly108, Phe134, Asp138, Ser139, Gly222, Gln223	Ala87, Ala106, Gly107, Gly108, Phe134, Asp138, Ser139, Gly222, Gln223
Pontes de Hidrogênio	MAN	Asp88, Ala106, Gly108, Gly222, Gln223, Gln224	Asp88, Ala106, Gly108, Gln223
	MMA	Asp88, Ala106, Gly108, Gly222, Gln223, Gln224	Ala87, Asp88, Ala106, Gly108, Gly222, Gln223, Gln224

Fonte: elaborada pelo autor.

5 CONCLUSÃO

Modelos confiáveis das proteínas ADL e AIL foram gerados com bons parâmetros em ferramentas de validação, permitindo o uso dessas estruturas como modelos representativos. Suas interações com carboidratos foram avaliadas por docking e dinâmica moleculares, o que revelou que essas proteínas interagem de forma diferente aos ligantes testados, apesar da similaridade estrutural dessas lectinas. O seu comportamento em solução aquosa foi observado pela captura de imagens realizada durante as simulações, ajudando a elucidar as propriedades de ligação e os resíduos envolvidos na ligação com carboidratos. Este estudo serve como um guia para pesquisas futuras sobre lectinas específicas de α -metil-D-manosídeo de *Arachis duranensis* e *Arachis ipaensis*.

6 REFERÊNCIAS

- ABRAHAM, M. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, v. 1-2, p. 19-25, 2015.
- ABRAHAM, M.; VAN DER SPOEL, D.; LINDAHL, E.; HESS, B. and the GROMACS development team, GROMACS User Manual version 5.1.2, www.gromacs.org, 2016.
- ADCOCK, S.; MCCAMMON, J. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chemical Reviews*, v. 106, n. 5, p. 1589-1615, 2006.
- AGRAWAL, B.B.L.; GOLDSTEIN, I. J. Specific Binding of Concanavalin A to Cross-Linked Dextran Gels, *Biochemical Journal*, v. 93, p. 23C-25C, 1965.
- AGRAWAL, P.; KUMAR, S.; DAS, H. Mass spectrometric characterization of isoform variants of peanut (*Arachis hypogaea*) stem lectin (SL-I). *Journal of Proteomics*, v. 73, n. 8, p. 1573-1586, 2010.
- ALENCAR, N. et al. Vatairea Macrocarpa Lectin Induces Paw Edema With Leukocyte Infiltration. *Protein & Peptide Letters*, v. 11, n. 2, p. 195-200, 2004.
- ALENCAR, N. M. et al. Leguminous lectins as tools for studying the role of sugar residues in leukocyte recruitment. *Mediators. Inflamm.*, v. 8, p. 107-113, 1999.
- ALENCAR, N. M. et al. The galactose binding lectin from Vatairea macrocarpa seeds induces in vivo neutrophil migration by indirect mechanism. *Int. J. Biochem. Cell Biol.*, v. 35, p. 1674-1681, 2003.
- ALMEIDA, A. C. et al. Purification and partial characterization of a new mannose/glucose-specific lectin from *Centrolobium tomentosum* Guill. ex Benth seeds exhibiting low toxicity on *Artemia* sp. *International Journal of Indigenous Medical Plants*, v. 47, p. 1567-1577, 2014.
- ALMEIDA, A. et al. Structural analysis of *Centrolobium tomentosum* seed lectin with inflammatory activity. *Archives of Biochemistry and Biophysics*, v. 596, p. 73-83, 2016.
- AQVIST, J.; LUZHKOVA, V.; BRANDSDAL, B. Ligand Binding Affinities from MD Simulations. *Accounts of Chemical Research*, v. 35, n. 6, p. 358-365, 2002.
- ARARIPE, D.A., et al. Partial characterization and immobilization in CNBr-activated Sepharose of a native lectin from *Platypodium elegans* seeds (PELa) and comparative study of

edematogenic effect with the recombinant form. *International Journal of Biological Macromolecules*, v. 102, p. 323-330, 2017.

ARGOS, P.; TSUKIHARA, T.; ROSSMANN, G. A structural comparison of concanavalin A and tomato bushy stunt virus protein, *Journal of Molecular Evolution*, v. 15, p. 169–179, 1980.

ARYA, S.; SALVE, A.; CHAUHAN, S. Peanuts as functional food: a review. *Journal of Food Science and Technology*, v. 53, n. 1, p. 31-41, 2015.

ASSREUY, A. et al. Vasodilator effects of Diocleinae lectins from the *Canavalia* genus. *Naunyn-Schmiedeberg's Archives of Pharmacology*, v. 380, n. 6, p. 509-521, 2009.

ASZODI, A.; TAYLOR, W. Homology modelling by distance geometry. *Folding and Design*, v. 1, n. 5, p. 325-334, 1996..

AUB, J.C.; SANFORD, B.H.; COTE, M.N. Studies on reactivity of tumor and normal cells to a wheat germ agglutinin. *Proceedings of the National Academy of Sciences of the United States*, v. 54, p. 396–399, 1965.

AUB, J.C.; TIESLAU, C.; LANKESTER, A. Reactions of normal and tumor cell to enzymes. I. Wheat-germ lipase and associated mucopolysaccharides. *Proceedings of the National Academy of Sciences of the United States*, v. 50, p. 613–619, 1963.

BANERJEE, R. et al. Conformation, Protein-Carbohydrate Interactions and a Novel Subunit Association in the Refined Structure of Peanut Lectin-Lactose Complex. *J. Mol. Biol.*, v. 259, p. 281-296, 1996.

BANERJEE, R. et al. Crystal structure of peanut lectin, a protein with an unusual quaternary structure. *Proceedings of the National Academy of Sciences of the United States of America*, v. 91, p. 227–231, 1994.

BATISTA DA NÓBREGA, R. et al. Structure of *Dioclea virgata* lectin: Relations between carbohydrate binding site and nitric oxide production. *Biochimie*, v. 94, n. 3, p. 900-906, 2012.

BENEVIDES, R. G. et al. A Lectin from *Platypodium elegans* with Unusual Specificity and Affinity for Asymmetric Complex N-Glycans. *Journal of Biological Chemistry*, v. 287, p. 26352-26364, 2012.

BENEVIDES, R.G. Avaliação do Potencial Fungicida e Termiticida de uma Fração Protéica Lectínica de Sementes de *Platypodium elegans* Vogel e Obtenção da Lectina Purificada.

Dissertação (Mestrado em Bioquímica) - Centro de Ciências, Universidade Federal do Ceará, Fortaleza, 151 p., 2008.

BENKERT, P.; BIASINI, M.; SCHWEDE, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, v. 27, n. 3, p. 343-350, 2010.

BENKERT, P.; TOSATTO, S.; SCHOMBURG, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*, v. 71, n. 1, p. 261-277, 2008.

BERTIOLI, D. et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nature Genetics*, v. 48, n. 4, p. 438-446, 2016.

BERWEGER, C.; VAN GUNSTEREN, W.; MÜLLER-PLATHE, F. Force field parametrization by weak coupling. Re-engineering SPC water. *Chemical Physics Letters*, v. 232, n. 5-6, p. 429-436, 1995.

BEZERRA, D. P. et al. Purificação de uma lectina presente em sementes de *Machaerium acutifolium*. *Anais da XXVI Reunião Nordestina de Botânica*, In: XXVI Reunião Nordestina de Botânica, Fortaleza, 2003.

BHAT, R. and VASHANTI, S. Occurrence of aflatoxins and its economic impact on human nutrition and animal feed. The new regulation. *Agricultural Development*, n. 23, p. 50-56, 1999.

BIASINI, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, v. 42, n. W1, p. W252-W258, 2014.

BLUNDELL, T. et al. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature*, v. 326, n. 6111, p. 347-352, 1987.

BÖHM, H. *Journal of Computer-Aided Molecular Design*, v. 12, n. 4, p. 309-309, 1998.

BOHM, H. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *Journal of Computer-Aided Molecular Design*, v. 6, n. 6, p. 593-606, 1992.

BOMPARD-GILLES, C. et al. Substrate mimicry in the active center of a mammalian α amylase: structural analysis of an enzyme-inhibitor complex. *Structure*, v. 4, n. 12, p. 1441-1452, 1996.

BOURNE, Y. et al. X-ray crystal structure determination and refinement at 1.9 Å resolution of isolectin I from the seeds of *Lathyrus ochrus*. *Journal of Molecular Biology*, v. 214, n. 2, p. 571-584, 1990.

BOYD, W. C. The proteins, Chemistry, Biological Activity, and Methods - the proteins of
BOYD, W. C.; REGUERA, R. M. Studies on haemagglutinins present in seeds of some representatives of the family Leguminosae. *Journal of Immunology*, v. 62, p. 333-339, 1949.

BRINT, A.; WILLETT, P. Algorithms for the Identification of Three-Dimensional Maximal Common Substructures. *Journal of Chemical Information and Computer Sciences*, v. 27, p. 152-158, 1987.

BROCKLEHURST, S.; PERHAM, R. Prediction of the three-dimensional structures of the biotinylated domain from yeast pyruvate carboxylase and of the lipoylated H-protein from the pea leaf glycine cleavage system: A new automated method for the prediction of protein tertiary structure. *Protein Science*, v. 2, n. 4, p. 626-639, 2008.

BROOKS, B. et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, v. 4, n. 2, p. 187-217, 1983.

BROWNE, W. et al. A possible three-dimensional structure of bovine alpha-lactalbumin based on that of hen's egg-white lysozyme. *Journal of Molecular Biology*, v. 42, n.1, p. 65-86, 1969.

BUTS, L. et al. Structural basis for the recognition of complex-type biantennary oligosaccharides by *Pterocarpus angolensis* lectin. *FEBS Journal*, v. 273, n. 11, p. 2407-2420, 2006.

CALVETE, J. et al. Molecular characterization and crystallization of Diocleinae lectins. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, v. 1430, n. 2, p. 367-375, 1999.

CALVETE, J. J. et al. Amino acid sequence, glycan structure, and proteolytic processing of the lectin of *Vatairea macrocarpa* seeds. *FEBS Letters*, v. 425, p. 286-292, 1998.

CARLSON, H.; JORGENSEN, W. An Extended Linear Response Method for Determining Free Energies of Hydration. *The Journal of Physical Chemistry*, v. 99, n. 26, p. 10667-10673, 1995.

CAVADA B.S. et al. Purification and characterization of a lectin from seeds of *Vatairea macrocarpa* Duke. *Phytochemistry*, v. 49, n. 3, p. 675-80, 1998.

CAVADA, B. et al. Structural studies and nociceptive activity of a native lectin from *Platypodium elegans* seeds (nPELa). *International Journal of Biological Macromolecules*, 2017.

CHARIFSON, P. et al. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *Journal of Medicinal Chemistry*, v. 42, n. 25, p. 5100-5109, 1999.

CHELVANAYAGAM, G; HERINGA, J.; ARGOS P. Anatomy and evolution of proteins displaying the viral capsid jellyroll topology. *Journal of Molecular Biology*, v. 228, p. 220–242, 1992.

CHEN, Y. Beware of docking!. *Trends in Pharmacological Sciences*, v. 36, n. 2, p. 78-95, 2015.

CLAESSENS, M. et al. Modelling the polypeptide backbone with 'spare parts' from known protein structures. *Protein Engineering*, v. 2, n. 5, p. 335-345, 1989.

CLARK, R.D. et al. Consensus scoring for ligand/protein interactions. *Journal of Molecular Graphics and Modelling*, v. 20, n. 4, p. 281–295, 2002.

CLAUSSEN, H. et al. FlexE: efficient molecular docking considering protein structure variations¹ Edited by J. Thornton. *Journal of Molecular Biology*, v. 308, n. 2, p. 377-395, 2001.

CORNELL, W. et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, v. 117, n. 19, p. 5179-5197, 1995.

DAM, T. K. et al. Binding studies of alpha-GalNAc-specific lectins to the alpha-GalNAc (Tn-antigen) form of porcine submaxillary mucin and its smaller fragments. *The Journal Of Biological Chemistry*., v. 282, p. 28256-28263, 2007.

DAMME, E. et al. Plant Lectins: A Composite of Several Distinct Families of Structurally and Evolutionary Related Proteins with Diverse Biological Roles. *Critical Reviews in Plant Sciences*, v. 17, n. 6, p. 575-692, 1998.

DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, v. 98, n. 12, p. 10089-10092, 1993.

DELANO, W.L. *The Pymol Molecular Graphics System*. DeLano Scientific, San Carlos, CA, 2002.

DELBAERE, L. et al. Structure of a Legume Lectin with an Ordered N-Linked Carbohydrate in Complex with Lactose, *Journal of Molecular Biology*, v. 230, p. 950–965, 1993.

DESJARLAIS, L. et al. Docking flexible ligands to macromolecular receptors by molecular shape. *Journal of Medicinal Chemistry*, v. 29, n.11, p. 2149–2153, 1986.

DESSEN, A. et al. X-ray Crystal Structure of the Soybean Agglutinin Cross-Linked with a Biantennary Analog of the Blood Group I Carbohydrate Antigen. *Biochemistry*, v. 34, n. 15, p. 4933-4942, 1995.

DHILLON, S.; RAKE, A.; MIKSCHE, J. Reassociation Kinetics and Cytophotometric Characterization of Peanut (*Arachis hypogaea* L.) DNA. *PLANT PHYSIOLOGY*, v. 65, n. 6, p. 1121-1127, 1980.

EDELMAN, G. et al. The Covalent and Three-Dimensional Structure of Concanavalin A. *Proceedings of the National Academy of Sciences of the United States of America*, v. 69, n. 9, p. 2580–2584, 1972.

EDELMAN, G.M.; WANG, J.L. Binding and Functional Properties of Concanavalin A and Its Derivatives. *Journal of Biological Chemistry*, v. 253, n. 9, p. 3016-3022, 1978

EINSPAH, H. et al. The Crystal Structure of Pea Lectin at 3.0-Å Resolution. *The Journal of Biological Chemistry*. v. 261, N. 35, p. 16518-16527, 1986

ELFSTRAND, M. *Über blutkörperchenagglutinierende Eiweisse*. In “Gorberdorfer Veröffentlichungen a. Band I” (R. Kobert, ed.), pp. 1–159 . Enke, Stuttgart, Germany, 1898.

EMSLEY, P. et al. Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography*, v. 66, n. 4, p. 486-501, 2010.

FEHER, M.; DERETAY, E.; ROY, S. BHB: A Simple Knowledge-Based Scoring Function to Improve the Efficiency of Database Screening. *Journal of Chemical Information and Computer Sciences*, v. 43, n. 4, p. 1316-1327, 2003.

- FISCHER, D. et al. Surface motifs by a computer vision technique: Searches, detection, and implications for protein-ligand recognition. *Proteins: Structure, Function, and Genetics*, v. 16, n. 3, p. 278-292, 1993.
- FISER, A.; ŠALI, A. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. *Methods in Enzymology*, p. 461-491, 2003.
- GALBRAITH, W.; GOLDSTEIN, I. Phytohemagglutinins: A new class of metalloproteins. Isolation, purification, and some properties of the lectin from *Phaseolus lunatus*. *FEBS Letters*, v. 9, n. 4, p. 197-201, 1970.
- GARCIA-PINO, A. et al. How a Plant Lectin Recognizes High Mannose Oligosaccharides. *PLANT PHYSIOLOGY*, v. 144, n. 4, p. 1733-1741, 2007.
- GASTEIGER, E. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, v. 31, n. 13, p. 3784-3788, 2003.
- GEGG C.V. et al. Characterization of the adenine binding sites of two *Dolichos biflorus* lectins. *Biochemistry*, v. 31, n. 30, p. 6938-6942, 1992.
- GEHLHAAR, D. et al. De Novo Design of Enzyme Inhibitors by Monte Carlo Ligand Generation. *Journal of Medicinal Chemistry*, v. 38, n. 3, p. 466-472, 1995.
- GEHLHAAR, D. et al. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chemistry & Biology*, v. 2, n. 5, p. 317-324, 1995.
- GEPTS, P. Legumes as a Model Plant Family. Genomics for Food and Feed Report of the Cross-Legume Advances through Genomics Conference. *PLANT PHYSIOLOGY*, v. 137, n. 4, p. 1228-1235, 2005.
- GOLDSTEIN, I. et al. What should be called a lectin? *Nature*, v. 285, p. 66 1980..
- GOLDSTEIN, I. J.; HOLLERMAN, C. E.; MERRICK, J. M. The interaction of Concanavalin A with polysaccharides. *Federation Proceedings*, v. 22, p. 297, 1963.
- GOLDSTEIN, I. J.; HOLLERMAN, C. E.; MERRICK, J. M. The interaction of polysaccharides with Concanavalin A. *Biochimica et Biophysica Acta*, v. 97, p. 68, 1965.

- GONÇALVES, F. M. et al. Vatairea macrocarpa Lectin (VML) Induces Depressive-like Behavior and Expression of Neuroinflammatory Markers in Mice. *Neurochemical Research*, v. 38, p. 2375-2384, 2013.
- GOODSELL, D. et al. Automated docking in crystallography: Analysis of the substrates of aconitase. *Proteins: Structure, Function, and Genetics*, v. 17, n. 1, p. 1-10, 1993.
- GRABIELE, M. et al. Genetic and geographic origin of domesticated peanut as evidenced by 5S rDNA and chloroplast DNA sequences. *Plant Systematics and Evolution*, v. 298, n. 6, p. 1151-1165, 2012.
- GREER, J. Comparative model-building of the mammalian serine proteases. *J. Journal of Molecular Biology*, v. 153, n. 4, p. 1027-1042, 1981.
- GU, J. et al. Potential roles of N-glycosylation in cell adhesion. *Glycoconjugate Journal*, v. 29, n. 8-9, p. 599-607, 2012.
- GU, J. et al. Potential roles of N-glycosylation in cell adhesion. *Glycoconjugate Journal*, v. 29, n. 8-9, p. 599-607, 2012.
- HAMELRYCK T.W. et al. Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. *Journal of Molecular Biology*, v. 286, n. 4, p. 1161-1177, 1999.
- HAMELRYCK, T. et al. Crystal Structure of Arcelin-5, a Lectin-like Defense Protein from *Phaseolus vulgaris*. *Journal of Biological Chemistry*, v. 271, n. 51, p. 32796-32802, 1996.
- HAMELRYCK, T. et al. The Crystallographic Structure of Phytohemagglutinin-L. *Journal of Biological Chemistry*, v. 271, n. 34, p. 20479-20485, 1996.
- HANSEN, J. E.; NIELSEN C. M.; NIELSEN, C.; NIELSEN, P.; HEEGAARD, L.; R. MATHIESEN, J.O. Nielsen, Correlation between carbohydrate structures on the envelope glycoprotein gp120 of HIV-1 and HIV-2 and syncytium inhibition with lectins, *J. Acquired Journal of Acquired Immune Deficiency Syndromes*. v.3, p. 635-641, 1989.
- HARDMAN, K.; AINSWORTH, C. Structure of concanavalin A at 2.4-Å resolution. *Biochemistry*, v. 11, n. 26, p. 4910-4919, 1972.
- HART, T.; READ, R. A multiple-start Monte Carlo docking method. *Proteins: Structure, Function, and Genetics*, v. 13, n. 3, p. 206-222, 1992.

- HAVEL, T.; SNOW, M. A new method for building protein conformations from sequence alignments with homologues of known structure. *Journal of Molecular Biology*, v. 217, n. 1, p. 1-7, 1991.
- HEAD, R. et al. VALIDATE: A New Method for the Receptor-Based Prediction of Binding Affinities of Novel Ligands. *Journal of the American Chemical Society*, v. 118, n. 16, p. 3959-3969, 1996.
- HESS, B. et al. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, v. 18, n. 12, p. 1463-1472, 1997.
- HOOVER, W. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*, v. 31, p. 1695-1697, 1985.
- HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, v. 14, n. 1, p. 33-38, 1996.
- ISHCHENKO A. and SHAKHNOVICH, E. Small Molecule Growth 2001 (SMoG2001): An Improved Knowledge-Based Scoring Function for Protein-Ligand Interactions. *Journal of Medicinal Chemistry*, n. 45, v. 13, p. 2770-278, 2002.
- JENNINGS, A.; EDGE, C.; STERNBERG, M. An approach to improving multiple alignments of protein sequences using predicted secondary structure. *Protein Engineering, Design and Selection*, v. 14, n. 4, p. 227-231, 2001.
- JIANG, F.; KIM, S. "Soft docking": Matching of molecular surface cubes. *Journal of Molecular Biology*, v. 219, n. 1, p. 79-102, 1991.
- JONES, G. et al. Development and validation of a genetic algorithm for flexible docking 1 Edited by F. E. Cohen. *Journal of Molecular Biology*, v. 267, n. 3, p. 727-748, 1997.
- JONES, T.; THIRUP S., Using known substructures in protein model building and crystallography. *The EMBO Journal*, v. 5, n. 4, p. 819-822, 1986.
- JOUBERT, F. J.; SHARON, N.; MERRIFIELD, E. H. Purification and properties of a lectin from *Lonchocarpus capassa* (apple-leaf) seed. *Phytochemistry*, v. 25, n. 2, p. 323-327, 1986.
- JUNG, E.; Brunak S. Prediction of N-glycosylation sites in human proteins, in preparation, 2004.

KANELLOPOULOS, P. et al. A Triclinic Crystal Form of the Lectin Concanavalin A. *Journal of Structural Biology*, v. 117, n. 1, p. 16-23, 1996.

KIESSLING, L.; POHL, N. Strength in numbers: non-natural polyvalent carbohydrate derivatives. *Chemistry & Biology*, v. 3, n. 2, p. 71-77, 1996.

KISS, R. et al. In vitro influence of *Phaseolus vulgaris*, *Griffonia simplicifolia*, concanavalin A, wheat germ, and peanut agglutinins on HCT-15, LoVo, and SW837 human colorectal cancer cell growth. *Gut*, v. 40, n. 2, p. 253-261, 1997.

KOCHERT, G. et al. RFLP and Cytogenetic Evidence on the Origin and Evolution of Allotetraploid Domesticated Peanut, *Arachis hypogaea* (Leguminosae). *American Journal of Botany*, v. 83, n. 10, p. 1282, 1996.

KOEHL, P.; LEVITT, M. *Nature Structural Biology*, v. 6, n. 2, p. 108-111, 1999.

KOLLMAN, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*, v. 93, n. 7, p. 2395-2417, 1993.

KORB, O.; STÜTZLE, T.; EXNER, T. Empirical Scoring Functions for Advanced Protein–Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling*, v. 49, n. 1, p. 84-96, 2009.

KOZIARA, K. et al. Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. *Journal of Computer-Aided Molecular Design*, v. 28, n. 3, p. 221-233, 2014.

KUNTZ, I.; LEACH A. Conformational analysis of flexible ligands in macromolecular receptor sites. *Journal of Computational Chemistry*, v. 13, p. 730–748, 1992.

LANDSTEINER, K. and Raubitschek, H. (1907). Beobachtungen über Hamolyse und Hamagglutination. *Zentralbl. Bakteriol. Parasitenk. Infektionskr. Hyg. Abt. 1: Orig.* 45, p. 660–667.

LASKOWSKI, R. et al. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, v. 26, n. 2, p. 283-291, 1993.

LEVITT, M. Accurate modeling of protein conformation by automatic segment matching. *Journal of Molecular Biology*, v. 226, n. 2, p. 507-533, 1992.

LEWIS, G. et al. *Legumes of the world*. Royal Botanical Gardens, Kew, 2005.

- LIGPLOT: Laskowski R A, Swindells M B (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.*, 51, 2778-2786.
- LOREA, P. et al. In vitro characterization of lectin-induced alterations on the proliferative activity of three human melanoma cell lines. *Melanoma Research*, v. 7, n. 5, p. 353-363, 1997.
- LORIS, R. et al. A Structure of the Complex between Concanavalin A and Methyl-3,6-di-O-(α -D-mannopyranosyl)- α -D-mannopyranoside Reveals Two Binding Modes. *Journal of Biological Chemistry*, v. 271, n. 48, p. 30614-30618, 1996.
- LORIS, R. et al. Crystal structure determination and refinement at 2.3- \AA resolution of the lentil lectin. *Biochemistry*, v. 32, n. 34, p. 8772-8781, 1993.
- LORIS, R. et al. Crystal structure of Pterocarpus angolensis lectin in complex with glucose, sucrose and turanose. *J. Biol. Chem.*, v. 278, p. 16297-16303, 2003.
- LORIS, R. et al. Legume lectin structure. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, v. 1383, n. 1, p. 9-36, 1998.
- LORIS, R. et al. Structural analysis of two crystal forms of lentil lectin at 1.8 \AA resolution. *Proteins: Structure, Function, and Genetics*, v. 20, n. 4, p. 330-346, 1994.
- LORIS, R. et al. Structural Basis of Oligomannose Recognition by the Pterocarpus angolensis Seed Lectin. *Journal of Molecular Biology*, v. 335, n. 5, p. 1227-1240, 2004.
- LOSSIO, C. et al. Lectin from Canavalia villosa seeds: A glucose/mannose-specific protein and a new tool for inflammation studies. *International Journal of Biological Macromolecules*, v. 105, p. 272-280, 2017.
- LOSSIO, C. F. et al. Isolation, purification and partial physicochemical characterization of a lectin in Andira Pisonis. Mart seed. *BMC Proc.*, v. 8, suppl. 4, p. 226, 2014.
- LOTAN, R. The purification, composition, and specificity of the anti-T lectin from peanut (*Arachis hypogaea*). *Journal of Biological Chemistry*, v. 250, n.21, p. 8518-8523, 1975.
- LÜTHY, R.; BOWIE, J.; EISENBERG, D. Assessment of protein models with three-dimensional profiles. *Nature*, v. 356, n. 6364, p. 83-85, 1992.
- MACKERELL, A. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *The Journal of Physical Chemistry B*, v. 102, n. 18, p. 3586-3616, 1998.

- MALDE, A. et al. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *Journal of Chemical Theory and Computation*, v. 7, n. 12, p. 4026-4037, 2011.
- MALIARIK, M. et al. ESR and fluorescence studies on the adenine binding site of lectins using a spin-labeled analog; *Biochemistry*, v. 28, n. 2, p. 912–917, 1989.
- MARONDEDZE, C.; CHIDZWONDO, F.; SITHOLE-NIANG, I.; READ, J. S. Pterocarpus rotundifolius lectin gene. Submitted to the EMBL/GenBank/DDBJ databases, 2004.
- MARQUES, G. et al. Contribution of the carbohydrate-binding ability of Vatairea guianensis lectin to induce edematogenic activity. *Biochimie*, v. 140, p. 58-65, 2017.
- MARTINS, A.M. et al. Renal effects induced by the lectin from Vatairea macrocarpa seeds. *J. Pharm. Pharmacol.*, v. 57, p. 1329-1333, 2005.
- MARTÍ-RENOM, M. et al. Comparative Protein Structure Modeling of Genes and Genomes. *Annual Review of Biophysics and Biomolecular Structure*, v. 29, n. 1, p. 291-325, 2000.
- MATSUMOTO, I.; OSAWA, T. Purification and characterization of an anti-H(O) phytohemagglutinin of Ulex europeus. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, v. 194, n. 1, p. 180-189, 1969.
- MCCUBBIN, W.; KAY, C. Molecular weight studies on concanavalin A. *Biochemical and Biophysical Research Communications*, v. 44, n. 1, p. 101-109, 1971.
- MENG, X. et al. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Current Computer Aided-Drug Design*, v. 7, n. 2, p. 146-157, 2011.
- MOREIRA, G. et al. Structure Predictions of Two Bauhinia variegata Lectins Reveal Patterns of C-Terminal Properties in Single Chain Legume Lectins. *PLoS ONE*, v. 8, n. 11, p. e81338, 2013.
- MORETZSOHN, M. et al. A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Annals of Botany*, v. 111, n. 1, p. 113-126, 2012.
- MORRIS, G. et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, v. 19, n. 14, p. 1639-1662, 1998.

- NAISMITH, J. et al. Refined structure of cadmium-substituted concanavalin A at 2.0 Å resolution. *Acta Crystallographica Section D Biological Crystallography*, v. 49, n. 6, p. 561-571, 1993.
- NAPIMOGA, M. H. et al. Lonchocarpus sericeus lectin decreases leukocyte migration and mechanical hypernociception by inhibiting cytokine and chemokines production. *Int. Immunopharmacol.*, v. 7, n. 6, p. 824-835, 2007.
- NASCIMENTO, K. S. et al. Purification of a thermostable antinociceptive lectin isolated from *Andira anthelmia*. *J. Mol. Recognit.*, 2015. doi: 10.1002/jmr.2523 (Epub ahead of print).
- NIELEN, S. et al. Matita, a new retroelement from peanut: characterization and evolutionary context in the light of the *Arachis A–B* genome divergence. *Molecular Genetics and Genomics*, v. 287, n. 1, p. 21-38, 2011.
- NJAPAU, H.; MUZUNGAILE, E.; CHANGA, R. The effect of village processing techniques on the content of aflatoxins in corn and peanuts in Zambia. *Journal of the Science of Food and Agriculture*, v. 76, n. 3, p. 450-456, 1998.
- NOBRE, C.B. Caracterização físico-química e efeito sobre bactérias orais de uma lectina de sementes de *Andira surinamensis* (Bondt) Splitg. ex Amshoff. Dissertação (Mestrado em Bioquímica) - Centro de Ciências, Universidade Federal do Ceará, Fortaleza, 2012.
- NOREL, R. et al. Molecular surface recognition by a computer vision-based technique. "Protein Engineering, Design and Selection", v. 7, n. 1, p. 39-46, 1994.
- NOSE, S. A unified formulation of the constant temperature molecular dynamics methods, *Journal of Chemical Physics*, v. 81, p. 511-519, 1984.
- NOWELL P.C. Phytohemagglutinin: An initiator of mitosis in cultures of normal human leukocytes. *Cancer Research*, v. 20, p. 462–466, 1960.
- OSHIRO, C.; KUNTZ, I.; DIXON, J. Flexible ligand docking using a genetic algorithm. *Journal of Computer-Aided Molecular Design*, v. 9, n. 2, p. 113-130, 1995.
- PARK, Y. et al. Lectin-based optical sensing for quantitative analysis of cancer antigen CA15-3 as a breast cancer marker. *Sensors and Actuators B: Chemical*, v. 186, p. 571-579, 2013.
- PARRINELLO, M.; RAHMAN, A. Polymorphic transitions in single crystals: a new molecular dynamics method, *Journal of Applied Physics*, v. 52, p. 7182-7190, 1981.

- PATHAK, M. et al. Molecular cloning, expression, and cytokinin (6-benzylaminopurine) antagonist activity of peanut (*Arachis hypogaea*) lectin SL-I. *Plant Molecular Biology*, v. 62, n. 4-5, p. 529-545, 2006.
- PEARSON, R. Effective protein sequence comparison. *Methods in Enzymology*, v. 266, p. 227-258, 1996.
- PEREIRA-JUNIOR, F. et al. Purification and characterization of a mannose/N-acetyl- d-glucosamine-specific lectin from the seeds of *Platymiscium floribundum* Vogel. *Journal of Molecular Recognition*, v. 25, n. 8, p. 443-449, 2012.
- PETERSEN, T. et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, v. 8, n. 10, p. 785-786, 2011.
- PETTERSEN, E. et al. UCSF Chimera: A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, v. 25, n. 13, p. 1605-1612, 2004.
- PEUMANS, W.; VAN DAMME, E. Lectins as Plant Defense Proteins. *Plant Physiology*, v. 109, n. 2, p. 347-352, 1995.
- PINTO-JUNIOR, V. et al. Crystal structure of *Pisum arvense* seed lectin (PAL) and characterization of its interaction with carbohydrates by molecular docking and dynamics. *Archives of Biochemistry and Biophysics*, v. 630, p. 27-37, 2017a.
- PINTO-JUNIOR, V. et al. Molecular modeling, docking and dynamics simulations of the *Dioclea lasiophylla* Mart. Ex Benth seed lectin: An edematogenic and hypernociceptive protein. *Biochimie*, v. 135, p. 126-136, 2017b.
- PINTO-JUNIOR, V. et al. Structural studies of a vasorelaxant lectin from *Dioclea reflexa* Hook seeds: Crystal structure, molecular docking and dynamics. *International Journal of Biological Macromolecules*, v. 98, p. 12-23, 2017c.
- PIRES, A. et al. Atividade Antiinflamatória Da Lectina De Sementes De *Lonchocarpus araripensis*. 3ª Reunião Regional FESBE, Fortaleza, CE, 2008.
- PIRES, A. F. et al. A novel N-acetyl-glucosamine lectin of *Lonchocarpus araripensis* attenuates acute cellular inflammation in mice. *Inflammation Research (Printed ed.)*, v. 65, p. 43-52, 2016.
- RAMOS, L. et al. Chromosomal and phylogenetic context for conglutin genes in *Arachis* based on genomic sequence. *Molecular Genetics and Genomics*. v. 275, p. 578–592, 2006.

- RANGEL, T. B. A. et al. Isolamento e caracterização parcial de lectina proveniente de sementes da espécie vegetal *Andira fraxinifolia* (Leguminosae-Fabaceae). *Anais/Resumos da 61ª Reunião Anual da SBPC*, issn 2176-1221, 2009.
- RAREY, M. et al. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *Journal of Molecular Biology*, v. 261, n. 3, p. 470-489, 1996.
- REISNER, Y.; LINKER-ISRAELI, M.; SHARON, N. Separation of mouse thymocytes into two subpopulations by the use of peanut agglutinin. *Cellular Immunology*, v. 25, n. 1, p. 129-134, 1976.
- ROBERT, X.; GOUET, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, v. 42, n. W1, p. W320-W324, 2014.
- ROBERTS, D.D.; GOLDSTEIN, I.J. Adenine Binding Sites of the Lectin from Lima Beans (*Phaseolus lunatus*). *Journal of Biological Chemistry*, v. 258, n. 22, p. 13820-13824, 1983
- ROY, K.; KAR, S.; DAS, R. Other Related Techniques. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, p. 357-425, 2015.
- SABATE, J.; ANG, Y. Nuts and health outcomes: new epidemiologic evidence. *American Journal of Clinical Nutrition*, v. 89, n. 5, p. 1643S-1648S, 2009.
- SALVEMINI, D. et al. Nitric oxide: a key mediator in the early and late phase of carrageenan-induced rat paw inflammation. *British Journal of Pharmacology*, v. 118, n. 4, p. 829-838, 1996.
- SAMOLUK, S. et al. First insight into divergence, representation and chromosome distribution of reverse transcriptase fragments from L1 retrotransposons in peanut and wild relative species. *Genetica*, v. 143, n. 1, p. 113-125, 2015.
- SANCHEZ R. et al. Protein structure modeling for structural genomics. *Nature Structural Biology*, v. 7, p. 986-990, 2000.
- SÁNCHEZ, R. et al. *Nature Structural Biology*, v. 7, p. 986-990, 2000.
- SANZ-APARICIO, J. et al. The crystal structure of *Canavalia brasiliensis* lectin suggests a correlation between its quaternary conformation and its distinct biological properties from Concanavalin A. *FEBS Letters*, v. 405, n. 1, p. 114-118, 1997.
- SCHMID, N. et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal*, v. 40, n. 7, p. 843-856, 2011.

SCHWARZ, F.; AEBI, M. Mechanisms and principles of N-linked protein glycosylation. *Current Opinion in Structural Biology*, v. 21, n. 5, p. 576-582, 2011.

SHAANAN, B.; LIS, H.; SHARON, N. Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science*, v. 254, n. 5033, p. 862-866, 1991.

SHARON, N.; LIS, H. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*, v. 14, n. 11, p. 53-62, 2004.

SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, v. 7, n. 1, p. 539-539, 2014.

SILVA, H. et al. Purification and primary structure determination of a galactose-specific lectin from *Vatairea guianensis* Aublet seeds that exhibits vasorelaxant effect. *Process Biochemistry*, v. 47, n. 12, p. 2347-2355, 2012.

SINGH, R.; DAS, H. Purification of lectins from the stems of peanut plants. *Glycoconjugate Journal*, v. 11, n. 4, p. 282-285, 1994.

SOUSA, B. L. et al. High-resolution structure of a new Tn antigen-binding lectin from *Vatairea macrocarpa* and a comparative analysis of Tn-binding legume lectins. *International Journal of Biochemistry & Cell Biology*, v. 59, p. 103-110, 2015.

SRINIVASAN, S.; MARCH, C.; SUDARSANAM, S. An automated method for modeling proteins on known templates using distance geometry. *Protein Science*, v. 2, n. 2, p. 277-289, 2008.

STANLEY, P. et al. Chapter 8: N-Glycans, in: A. Varki, R.D. Cummings, J.D. Esko, et al. (Eds.), *Essentials of Glycobiology*, second ed., Cold Spring Harbor, New York, 2009.

SUMNER, J.B. and HOWELL, S.F. The identification of the hemagglutinin of the jack bean with concanavalin *American Journal of Bacteriology*, v. 32, p. 227-237, 1936.

TAKAHASHI, T.; RAMACHANDRAMURTHY, P.; LIENER, I. Some physical and chemical properties of a phytohemagglutinin isolated from *Phaseolus vulgaris*. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, v. 133, n. 1, p. 123-133, 1967.

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

VAN DAMME, E. et al. The seed lectins of black locust (*robinia pseudoacacia*) are encoded by two genes which differ from the bark lectin genes. *Plant Molecular Biology*, v. 29, n. 6, p. 1197-1210, 1995.

VAN DAMME, E. J. M. et al. Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Critical Review in Plant Sciences* v. 17, p. 575–692, 1998.

VASCONCELOS, M. A. et al. Purification and primary structure of a novel mannose-specific lectin from *Centrolobium microchaete* Mart seeds. *International Journal of Biological Macromolecules*, v. 81, p. 600-607, 2015.

VERKHIVKER, G. et al. *Journal of Computer-Aided Molecular Design*, v. 14, n. 8, p. 731-751, 2000.

WATKINS, W. M.; MORGAN, W. T. J. Neutralization of the anti-H agglutinin in eel by simple sugars. *Nature* v. 169, p. 825–826, 1952.

WEINER, S. et al. A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, v. 106, n. 3, p. 765-784, 1984.

YANG, D.C.H. et al. Rotational Correlation Time of Concanavalin A after Interaction with a Fluorescent Probe. *Journal of Biological Chemistry*, v. 249, n. 21, p. 7018-7023, 1974.

7 ANEXO I

TRABALHOS PUBLICADOS DURANTE A GRADUAÇÃO






International Journal of Biological Macromolecules

Volume 102, September 2017, Pages 323-330



Partial characterization and immobilization in CNBr-activated Sepharose of a native lectin from *Platypodium elegans* seeds (PELa) and comparative study of edematogenic effect with the recombinant form

David Alencar Araripe ^a, Vanir Reis Pinto-Junior ^a, Antonio Hadson Bastos Neco ^a, Mayara Queiroz Santiago ^a, Vinicius Jose Silva Osterne ^a, Alana Freitas Pires ^b, Claudia Figueiredo Lossio ^a, Maria Gleiciane Queiroz Martins ^a, Jorge Luiz Almeida Correia ^a, Raquel Guimaraes Benevides ^c, Rodrigo Bainy Leal ^d, Ana Maria Sampaio Assreuy ^b, Kyria Santiago Nascimento ^a , Benildo Sousa Cavada ^a  








International Journal of Biological Macromolecules

Available online 1 September 2017

In Press, Corrected Proof 

Structural studies and nociceptive activity of a native lectin from *Platypodium elegans* seeds (nPELa)

Benildo Sousa Cavada ^a  , David Alencar Araripe ^a, Ivanice Bezerra Silva ^a, Vanir Reis Pinto-Junior ^a, Vinicius Jose Silva Osterne ^a, Antonio Hadson Bastos Neco ^a, Eva Pollyanna Peixe Laranjeira ^b, Claudia Figueiredo Lossio ^a, Jorge Luis Almeida Correia ^a, Alana Freitas Pires ^b, Ana Maria Sampaio Assreuy ^b  , Kyria Santiago Nascimento ^a 

8 ANEXO II

TRABALHO DA MONOGRAFIA SUBMETIDO

REVISTA:

*Journal of Molecular Modelling***Homology modelling, molecular docking and dynamics of two α -methyl-D-mannoside-specific lectins from *Arachis* genus**

Kyria Santiago Nascimento¹, David Alencar Araripe¹, Vanir Reis Pinto-Júnior¹,
Vinicius José Silva Osterne¹, Francisco William Viana Martins¹, Antônio Hadson
Bastos Neco¹, Gil Aquino Farias², Benildo Sousa Cavada^{1*}

Department of Biochemistry and Molecular Biology, Federal University of Ceara¹

Department of Physics, Federal University of Ceara²

*Corresponding author (email address: bscavada@ufc.br) Fax/Phone number: +55 85 33669818