



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOQUÍMICA**

**PAULO ABRAÃO CAVALCANTE MARANHÃO**

**ANÁLISE DE UMA SEQUÊNCIA DE cDNA OBTIDA DAS SEMENTES DE *Swartzia*  
*grandiflora* Bong. Benth QUE CODIFICA UMA LECTINA**

**FORTALEZA**

**2015**

PAULO ABRAÃO CAVALCANTE MARANHÃO

ANÁLISE DE UMA SEQUÊNCIA DE cDNA OBTIDA DAS SEMENTES DE *Swartzia grandiflora* Bong. Benth QUE CODIFICA UMA LECTINA

Dissertação apresentada ao Programa de Pós-Graduação em Bioquímica, do Departamento de Bioquímica e Biologia Molecular, da Universidade Federal do Ceará, como requisito parcial para a obtenção do Título de Mestre em Bioquímica, Área de Concentração: Bioquímica Vegetal.

Orientador: Prof. Dr. Thalles Barbosa Grangeiro.

FORTALEZA

2015

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- M26a Maranhão, Paulo Abraão Cavalcante.  
Análise de uma sequênciade cDNA obtida das sementes de *Swartzia grandiflora* Bong. Benth que codifica uma lectina / Paulo Abraão Cavalcante Maranhão. – 2015.  
101 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Bioquímica, Fortaleza, 2015.  
Orientação: Prof. Dr. Thalles Barbosa Grangeiro.
1. Biologia molecular vegetal. I. Título.

CDD 572

---

PAULO ABRAÃO CAVALCANTE MARANHÃO

ANÁLISE DE UMA SEQUÊNCIA DE cDNA OBTIDA DAS SEMENTES DE *Swartzia grandiflora* Bong. Benth QUE CODIFICA UMA LECTINA

Dissertação apresentada ao Programa de Pós-Graduação em Bioquímica, do Departamento de Bioquímica e Biologia Molecular, da Universidade Federal do Ceará, como requisito parcial para a obtenção do Título de Mestre em Bioquímica, Área de Concentração: Bioquímica Vegetal.

Aprovada em: \_\_\_/\_\_\_/\_\_\_\_\_.

BANCA EXAMINADORA

---

Prof. Dr. Thalles Barbosa Grangeiro  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Bruno Anderson Matias da Rocha  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Renato de Azevedo Moreira  
Universidade de Fortaleza (UNIFOR)

A Deus, pelo seu amor.

A minha família, pelo apoio integral.

Aos amigos, pela contribuição.

A esses, dedico com orgulho.

## AGRADECIMENTOS

Ao Prof. Dr. Thalles Barbosa Grangeiro, por me aceitar em seu grupo de pesquisa e dedicar parte de seu tempo para a minha orientação.

Ao Prof. Dr. José Francisco de Carvalho Gonçalves e à Dra. Andreia Varmes Fernandes, pelo incentivo e contribuição durante todo esse tempo em que nos conhecemos.

Ao Prof. Dr. Márcio Viana Ramos, pela ajuda, pelas dicas e pelas indagações que levaram à finalização desta pesquisa.

Ao Prof. Dr. Bruno Anderson Matias da Rocha, pela sua prontidão na elaboração desta pesquisa.

À Profa. Dra. Ilka Maria Vasconcelos, pela ajuda dada em parte das análises aqui apresentadas.

Ao Dr. Claudener Souza Teixeira, pela contribuição nas análises de bioinformática e por todo o auxílio oferecido.

Ao Dr. Frederico Bruno Mendes Batista Moreno e à Dra. Marina Duarte Pinto Lobo, pelas análises prestadas.

A todos os integrantes do Laboratório de Biologia Molecular, pela amizade, companheirismo, brincadeiras e delírios.

A todos aqueles que, direta ou indiretamente, contribuíram para a realização deste trabalho: Muito obrigado!

“Não há assunto tão velho que não possa ser dito algo de novo sobre ele”. (Fiodor Dostoievski)

## RESUMO

*Swartzia grandiflora* é uma leguminosa da tribo *Swartzieae*, um táxon basal da subfamília *Papilionoideae*, ainda pouco estudada em relação à ocorrência e caracterização de lectinas. O objetivo principal do presente trabalho foi obter sequências de cDNA, codificando lectinas de *S. grandiflora* e caracterizar as estruturas codificadas. Como resultado, foram obtidas seis sequências únicas de cDNA, codificando seis cadeias polipeptídicas distintas, denominadas de SGL-A a SGL-F. As estruturas primárias das proteínas deduzidas das sequências de cDNA possuíam 240 (SGL-A a SGL-D) e 244 (SGL-E e SGL-F) resíduos de aminoácidos. A massa molecular calculada variou de ~26.6 (SGL-A) a ~26.9 kDa (SGL-F), enquanto que o pI predito a partir das sequências de aminoácidos variou de 5.22 (SGL-B) a 5.37 (SGL-A, -C e -D). Quando as sequências de aminoácidos foram alinhadas e comparadas, entre si, por meio de uma matriz de identidade, foi possível agrupá-las em dois grupos, contendo 4 (de SGL-A a SGL-D; identidade média = 99,1%) e 2 (SGL-E e SGL-F; identidade média = 99,5%) sequências, respectivamente. A identidade média entre as sequências dos dois grupos foi 53,2%. Considerando a grande identidade entre as sequências de aminoácidos de um mesmo grupo, uma sequência representativa de cada um deles foi escolhida (SGL-A e SGL-E) para análises posteriores. Buscas por sequências homólogas no banco de proteínas do NCBI revelaram que SGL-A e SGL-E apresentam similaridade com lectinas de Fabaceae, porém com percentuais de identidade nunca superiores a 50-51%. Isso sugere que as lectinas de *S. grandiflora* identificadas na presente investigação, constituem um grupo distinto dentre as lectinas de leguminosas caracterizadas até o momento. Comparações com o banco de proteínas com estruturas tridimensionais conhecidas (PDB), demonstraram uma maior identidade de SGL-A (Bit score = 196, E-value =  $7e-61$ , identidade = 44%) com uma lectina (PELa) de *Platypodiumelegans* (tribo *Dalbergieae*) e de SGL-E (Bit score = 179, E-value =  $1e-54$ , identidade = 47%) com a aglutinina (BMA) de *Leucomphalos (Bowringia) mildbraedii* (tribo *Sophoreae*). O alinhamento das sequências de aminoácidos de SGL-A e SGL-E com essas estruturas permitiu que os resíduos de aminoácidos, hipoteticamente envolvidos na ligação a cátions divalentes e na formação do sítio de interação com carboidratos, fossem mapeados nas respectivas estruturas primárias. Modelos das estruturas tridimensionais de SGL-A e SGL-E foram gerados por modelagem comparativa, usando o método M4T, e validados em relação a parâmetros estereoquímicos. Os modelos estruturais de ambas as proteínas apresentaram um dobramento em barril beta, compreendendo duas folhas antiparalelas, uma com sete fitas e outra com cinco fitas. Este dobramento é característico das



lectinas de leguminosas. Docking molecular sítio dirigido sugeriu que D-Galactose é um ligante provável de SGL-A, a julgar pela energia de interação total favorável ( $-76,2$  kcal/mol) estimada para o complexo proteína-ligante.

**Palavras-chave:** *Swartzia*. *Fabaceae*. Clonagem. Docking molecular.

## ABSTRACT

*Swartzia grandiflora* is a legume *Swartzieae* tribe, a basal taxon of *Papilionoideae* subfamily, poorly studied in relation to the occurrence and characterization of lectins. The main objective of this study was to obtain cDNA sequences encoding *S. grandiflora* lectins and characterize the encoded structures. As a result, there were obtained six unique cDNA sequences, encoding six distinct polypeptide chains, termed SGL-A and SGL-F. The primary structures of the proteins deduced from the cDNA sequence had 240 (SGL-A to D) and 244 (SGL-E and -F) amino acid residues. The calculated molecular weight ranged from ~ 26.6 (SGL-A) to ~ 26.9 kDa (SGL-F), whilst the pI predicted from the amino acid sequences varied from 5.22 (SGL-B) to 5.37 (SGL-A,- C and D). When the amino acid sequences were aligned and compared to each other by means of an identity matrix, it was possible to group them into two groups containing 4 (SGL-A to SGL-D, identity = 99.1%) and 2 ( SGL-E and SGL-F; identity = 99.5%) sequences, respectively. The average identity between the sequences of two groups was 53.2%. Considering the great identity between the amino acid sequences of a same group a representative sequence of each was chosen (SGL-A and SGL-E) for further analysis. Search for homologous sequences in the NCBI protein database revealed that SGL-A and SGL-E have similarity with lectins Fabaceae, but with ever higher percentage of identity to 50-51%. This suggests that lectins of *S. grandiflora* identified in this study constitute a distinct group from the legume lectins characterized to date. Comparisons with the bank of proteins with known three-dimensional structures (PDB) showed greater SGL-A identity (Bit score = 196, E-value = 7e-61, ID = 44%) with a lectin from *Platypodiumelegans* (tribe *Dalbergieae*) and SGL-E (Bit score = 179, E-value = 1e-54 = 47% identity) with the agglutinin (BMA) of *Leucomphalos (Bowringia) mildbraedii* (*Sophoreae* tribe). The alignment of the amino acids sequences SGL-A and SGL-E with such structures has allowed the hypothetical amino acid residues involved in binding divalent cations and the formation of the interaction site with carbohydrates were mapped in their primary structures. Models of three-dimensional structures of SGL-A and SGL-E were generated by comparative modeling using M4T method, and validated against stereo chemical parameters. Structural models of both proteins presented in a folding beta barrel, comprising two antiparallel sheets, one with seven tapes and other tapes of five. This folding is characteristic of legume lectins. Molecular docking site directed suggested that D-galactose is a likely binder SGL-A, judging by the favorable total interaction energy (-76.2 kcal / mol) estimated for the protein-ligand complex.

**Keywords:** *Swartzieae*. *Fabaceae*. Cloning. Molecular docking.

## LISTA DE FIGURAS

Figura 1 –	Árvore filogenética das tribos basais da subfamília <i>Papilionoideae</i> .....	24
Figura 2 –	Exemplificação do processo de interação entre lectinas e carboidratos da superfície celular.....	25
Figura 3 –	Ramo de <i>Swartzia grandiflora</i> com botão floral em desenvolvimento (A), ramo de <i>Swartzia grandiflora</i> com flor desenvolvida (B) e semente de <i>Swartzia grandiflora</i> .....	31
Figura 4 –	Representação do desenho dos iniciadores SGLForward e SGLReverse.....	34
Figura 5 –	Representação esquemática da extração de RNA das sementes de <i>Swartzia grandiflora</i> , utilizando a metodologia de precipitação por Cloreto de Lítio 10M.....	35
Figura 6 –	Eletroforese em gel de poliacrilamida 15 % da farinha de sementes de <i>Swartzia grandiflora</i> . A letra M indica o poço contendo o marcador molecular Low Molecular Weight Calibration Kit for SDS Electrophoresis (GE Healthcare). A letra A indica o poço onde foi aplicado 20 µg da farinha de <i>Swartzia grandiflora</i> , diluída em tampão da amostra, contendo β - mercaptoetanol 1 % .....	48
Figura 7 –	Identificação dos domínios conservados das sequências pelo banco de dados do NCBI. O domínio encontrado para cada uma das sequências está descrito na tarja azul. O valor +1 indica o <i>frame</i> de leitura da sequência.....	51
Figura 8 –	Identificação dos domínios conservados das sequências pelo banco de dados do NCBI. O domínio encontrado para cada uma das sequências está descrito na tarja azul. O valor +1 indica o <i>frame</i> de leitura da sequência.....	52
Figura 9 –	Alinhamento múltiplo das sequências nucleotídicas da RACE 3' que codificam para lectinas da <i>Swartzia grandiflora</i> . O número para o último nucleotídeo encontra-se descrito na margem direita do alinhamento. Os resíduos de aminoácidos conservados nas	

	sequências alinhadas estão demarcados com um asterisco (*).....	54
Figura 10 –	Alinhamento múltiplo das sequências nucleotídicas da RACE 3' que codificam para lectinas da <i>Swartzia grandiflora</i> . O número para o último nucleotídeo encontra-se descrito na margem direita do alinhamento. Os nucleotídeos conservados nas sequências alinhadas estão demarcados com um asterisco (*).....	55
Figura 11 –	Sequência SGL-A, codificando uma lectina de <i>Swartzia grandiflora</i> . O iniciador utilizado na RACE 3' encontra-se sublinhado. O códon de parada está indicado por um asterisco (*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente.....	56
Figura 12 –	Sequência SGL-B, codificando uma lectina de <i>Swartzia grandiflora</i> . O iniciador utilizado na RACE 3' encontra-se sublinhado. O códon de parada está indicado por um asterisco (*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente.....	57
Figura 13 –	Sequência SGL-C, codificando uma lectina de <i>Swartzia grandiflora</i> . O iniciador utilizado na RACE 3' encontra-se sublinhado. O códon de parada está indicado por um asterisco (*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente.....	58
Figura 14 –	Sequência SGL-D, codificando uma lectina de <i>Swartzia grandiflora</i> . O iniciador utilizado na RACE 3' encontra-se sublinhado. O códon de parada está indicado por um asterisco (*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último	

	resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente.....	59
Figura 15 –	Sequência SGL-E, codificando uma lectina de <i>Swartzia grandiflora</i> . O iniciador utilizado na RACE 3' encontra-se sublinhado. O códon de parada está indicado por um asterisco (*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente.....	60
Figura 16 –	Sequência SGL-F, codificando uma lectina de <i>Swartzia grandiflora</i> . O iniciador utilizado na RACE 3' encontra-se sublinhado. O códon de parada está indicado por um asterisco (*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente.....	61
Figura 17 –	Alinhamento múltiplo das sequências de aminoácidos que codificam para lectinas da <i>Swartzia grandiflora</i> . O número para o último resíduo de aminoácido encontra-se descrito na margem direita do alinhamento. Os resíduos de aminoácidos conservados nas sequências alinhadas estão demarcados com um asterisco (*)......	62
Figura 18 –	Predição da estrutura secundária da SGL-A, com a região N-terminal inserida pelo programa ProFunc.....	70
Figura 19 –	Predição da estrutura secundária da SGL-E, com a região N-terminal inserida pelo programa ProFunc.....	71
Figura 20 –	Predição topológica da estrutura secundária da SGL-A e SGL-E, com a região N-terminal inserida pelo programa ProFunc.....	72
Figura 21 –	Modelagem tridimensional das sequências deduzidas da SGL-A e SGL-E pelo programa M4T.....	73
Figura 22 –	Gráfico de Ramachandran, gerado pelo programa MolProbity, para validação da modelagem tridimensional da SGL-	

	A.....	74
Figura 23 –	Gráfico de Ramachandran, gerado pelo programa MolProbity, para validação da modelagem tridimensional da SGL-E.....	75
Figura 24 –	Resíduos de aminoácidos da SGL-A e SGL-E, possivelmente envolvidos na interação com Ca <sup>2+</sup> e Mn <sup>2+</sup> pelo alinhamento com outras lectinas de leguminosas presentes no NCBI.....	76
Figura 25 –	Sítios de interações preditos pelo CDD para a SGL-A. O resíduo de cor verde representa o sítio de N-glicosilação; os resíduos de cor azul representam os sítios de interação homodimérica e os resíduos de cor amarela representam os sítios de interações homotetraméricas.....	77
Figura 26 –	Sítios de interações preditos pelo CDD para a SGL-E. O resíduo de cor verde representa o sítio de N-glicosilação; os resíduos de cor azul representam os sítios de interação homodimérica e os resíduos de cor amarela representam os sítios de interações homotetraméricas.....	78
Figura 27 –	Interação entre Galactose e o Domínio de Reconhecimento a Carboidrato da SGL-A, obtida por docking molecular.....	80
Figura 28 –	Interação entre N-Acetil-D-Galactosamina e o Domínio de Reconhecimento a Carboidrato da SGL-A, obtida por docking molecular.....	81
Figura 29 –	Distância entre os resíduos de aminoácidos da SGL-A que realizam a interação com a galactose realizada pelo software ProFunc.....	82
Figura 30 –	Distância entre os resíduos de aminoácidos da SGL-E que realizam a interação com a galactose realizada pelo software ProFunc.....	83

## LISTA DE TABELAS

Tabela 1 –	Aplicações biotecnológicas relatadas na literatura científica para algumas lectinas.....	28
Tabela 2 –	Componentes utilizados na amplificação da região 5' UTR da SgL por <i>Nested</i> PCR.....	39
Tabela 3 –	Condições utilizadas para a amplificação das regiões 5' UTR e 3' UTR SgL por PCR.....	39
Tabela 4 –	Componentes utilizados na amplificação da região 3' UTR da SgL por PCR.....	40
Tabela 5 –	Sequências de aminoácidos com maior similaridade à região N-terminal da lectina de <i>Swartzia grandiflora</i> pelo programa BLASTp do NCBI, utilizando as sequências de proteínas não-redundantes como banco de dados (último acesso em 29 de agosto de 2015).....	49
Tabela 6 –	Matriz de diferença realizada pelo programa <i>BioEdit</i> das sequências nucleotídicas que codificam lectinas da semente de <i>S. grandiflora</i> .....	53
Tabela 7 –	Peso molecular e ponto isoelétrico estimado para as sequências de aminoácidos deduzidas, obtidos da RACE 3' da <i>S. grandiflora</i> .....	63
Tabela 8 –	Matriz de identidade (acima da diagonal) e matriz de diferença (abaixo da diagonal) para cada par de sequências comparadas. Estes números foram calculados com base no alinhamento múltiplo de sequências mostrado na figura 15.....	64
Tabela 9 –	Sequência de aminoácidos com maior similaridade à estrutura primária da SGL-A (bit score > 200) pelo programa BLASTp do NCBI, utilizando as sequências de proteínas não-redundantes como banco de dados (último acesso em 20 de agosto de 2015).....	65
Tabela 10 –	Sequência de aminoácidos com maior similaridade à estrutura primária da SGL-E (bit score > 186) pelo programa BLASTp do	

	NCBI, utilizando as sequências de proteínas não-redundantes como banco de dados (último acesso em 20 de agosto de 2015).....	66
Tabela 11 –	Lectinas com estrutura tridimensional conhecida e com maior similaridade à estrutura primária da SGL-A (Bit score > 166), obtido pelo BLASTp, utilizando o banco Protein Data Bank (PDB) (último acesso em 20 de agosto de 2015).....	67
Tabela 12 –	Lectinas com estrutura tridimensional conhecida e com maior similaridade à estrutura primária da SGL-E (Bit score > 160), obtido pelo BLASTp utilizando o banco Protein Data Bank (PDB) (último acesso em 20 de agosto de 2015).....	68



## LISTA DE ABREVIATURAS E SIGLAS

BLAST	Ferramenta Básica de Alinhamento Localizado
CIP	Fosfatase Alcalina do Intestino do Bezerro
DEPC	Dietil-pirocarbonato
DNA	Ácido desoxirribonucleico
DNase	Enzima endonuclease não específica que degrada DNA
dNTPs	Desoxirribonucleotídeo trifosfatados
<i>EcoRI</i>	Enzima de restrição
EDTA	Ácido etilenodiaminotetracético
GalNac	<i>N</i> -acetil-D-galactosamina
IPTG	Isopropil- $\beta$ -Galactosídeo
kb	Quilobase
LB	Meio de cultura Luria Bertani
min	Minuto
mM	Milimolar
M-MLV	Vírus da Leucemia de Murinos
pb	Pares de base
pGEM-T	Vetor plasmidial destinado à clonagem de produtos de PCR
PVP	Polivinilpirrolidona
RACE-PCR	Rápida Amplificação dos Finais do cDNA acoplado à reação em cadeia da polimerase
RNA	Ácido ribonucleico
RNase	Enzima ribonuclease
rpm	Rotações por minuto
RT	Transcrição Reversa
SDS	Dodecil Sulfato de Sódio
SOB	Meio de cultura para bactérias, suplementado com 2% de Triptona, 5% de Extrato de Levedura e 0,5% de NaCl pH 7,0
SOC	Meio super ótimo com repressão de catabólitos
TAP	Pirofosfatase Ácida do Tabaco
Tris	2-amino-2hidroximetilpropano-1,3-diol

$\mu\text{g}$	Micrograma
$\mu\text{L}$	Microlitro
U/ $\mu\text{L}$	Unidades de enzima por microlitros
UTR	Unidade de Término de Replicação

## SUMÁRIO

1	INTRODUÇÃO.....	20
2	FAMÍLIA FABACEAE.....	22
2.1	Subfamília <i>Caesalpinioideae</i> .....	22
2.2	Subfamília <i>Papilionoideae</i> .....	22
2.3	Subfamília <i>Mimosoideae</i> .....	24
3	LECTINAS.....	24
3.1	Breve Histórico da Lectina.....	26
3.2	Lectinas de <i>Fabaceae</i> .....	26
3.3	Classificação das Lectinas de <i>Fabaceae</i> .....	27
3.4	Função.....	27
3.5	Características moleculares.....	28
3.6	Clonagem molecular.....	29
3.7	Lectinas da tribo Swartzieae.....	29
4	OBJETIVOS.....	30
4.1	Objetivo geral.....	30
4.2	Objetivos específicos.....	30
5	MATERIAL.....	30
5.1	Sementes de <i>Swartzia grandiflora</i> .....	31
5.2	Microrganismos.....	31
5.3	Plasmídeos.....	32
5.4	Enzimas.....	32
5.5	Oligonucleotídeos.....	32
5.6	<i>Kits</i> e reagentes.....	32
6	MÉTODOS.....	33
6.1	Desenho e obtenção dos iniciadores.....	33
6.2	Maceração das sementes de <i>Swartzia grandiflora</i> .....	34
6.3	Extração de RNA total.....	34
6.4	Análise da integridade do RNA total.....	36
6.5	Tratamento do RNA com a DNase.....	36
6.6	Amplificação rápida das extremidades do cDNA (RACE).....	37
6.6.1	Amplificação da região 5' UTR.....	37

6.6.1.1 Tratamento do RNA total com a enzima CIP.....	37
6.6.1.2 Tratamento do RNA total com TAP.....	37
6.6.1.3 Reação de ligação do RNA tratado ao iniciador 5' RACE Adapter.....	38
6.6.1.4 Transcrição reversa do material ligado ao iniciador 5' RACE Adapter.....	38
6.6.1.5 Reação Nested PCR para obtenção da região 5' UTR.....	38
<b>6.6.2 Amplificação da região 3' UTR.....</b>	<b>39</b>
6.6.2.1 Transcrição reversa do RNA de <i>S. grandiflora</i> para a obtenção da região 3' UTR.....	39
6.6.2.2 Race 3'.....	40
<b>6.7 Clonagem.....</b>	<b>41</b>
<b>6.8 Ligação.....</b>	<b>41</b>
<b>6.9 Transformação.....</b>	<b>41</b>
<b>6.10 Extração de DNA plasmidial.....</b>	<b>42</b>
<b>6.11 Digestão enzimática com PvuII.....</b>	<b>43</b>
<b>6.12 Sequenciamento nucleotídico.....</b>	<b>43</b>
<b>6.13 Eletroforese em gel de poliacrilamida.....</b>	<b>43</b>
<b>6.14 Sequenciamento da região N – terminal.....</b>	<b>44</b>
<b>6.15 Espectrometria de massas.....</b>	<b>44</b>
<b>6.16 Análises de bioinformática.....</b>	<b>45</b>
<b>6.17 Modelagem computacional.....</b>	<b>46</b>
<b>6.18 Docking molecular.....</b>	<b>46</b>
<b>7 RESULTADOS.....</b>	<b>47</b>
7.1 Extração do RNA total das sementes de <i>Swartzia grandiflora</i> .....	47
7.2 Amplificação rápida das extremidades do cDNA (RACE).....	47
7.3 Caracterização in silico da região codificadora da SGL.....	50
<b>8 DISCUSSÃO.....</b>	<b>83</b>
<b>9 CONCLUSÃO.....</b>	<b>88</b>
<b>REFERÊNCIAS.....</b>	<b>89</b>
<b>APÊNDICE A – ELETROFORESE EM GEL DE AGAROSE 1,5% DO RNA EXTRAÍDO DAS SEMENTES DE SWARTZIA GRANDIFLORA.....</b>	<b>94</b>
<b>APÊNDICE B – ELETROFORESE EM GEL DE AGAROSE 1% DO PRODUTO DA RACE 3' DO RNA EXTRAÍDO DE SWARTZIA GRANDIFLORA.....</b>	<b>95</b>

<b>APÊNDICE C – ELETROFORESE EM GEL DE AGAROSE 1% DO PRODUTO DA RACE 3' DO RNA EXTRAÍDO DE <i>SWARTZIA GRANDIFLORA</i> (A) E DESSE MESMO PRODUTO LIGADO AO PGEM - T(B).....</b>	<b>96</b>
<b>APÊNDICE D – ELETROFORESE EM GEL DE AGAROSE 1% DO PRODUTO DA DIGESTÃO COM <i>PVUII</i> DOS CLONES TRANSFORMADOS COM OS PRODUTOS DA RACE 3' DO RNA EXTRAÍDO DE <i>SWARTZIA GRANDIFLORA</i>.....</b>	<b>97</b>
<b>APÊNDICE E – FONTES DE AUXÍLIO E FINANCIAMENTO.....</b>	<b>98</b>
<b>ANEXO A – MAPA DO VETOR DE CLONAGEM PGEM -T, DESTACANDO OS DIVERSOS SÍTIOS DE RESTRIÇÃO DE DIFERENTES ENZIMAS, A REGIÃO DO GENE QUE CONFERE RESISTÊNCIA À AMPICILINA E O GENE <i>LACZ</i>.....</b>	<b>99</b>

## 1 INTRODUÇÃO

A prospecção de diferentes biomoléculas é fundamental para a integridade dos ecossistemas nacionais, para as ciências básicas e para o fornecimento de produtos biotecnológicos para empresas especializadas (COUTINHO, 2006). Neste sentido, a obtenção de proteínas é de grande proveito, devido a sua ampla utilização em diferentes segmentos da sociedade (FERNANDES *et al.*, 2011). Porém, nem sempre, a purificação de determinados grupos de proteínas ocorre de uma maneira fácil. Deste modo, são necessários estudos mais específicos no sentido de facilitar a compreensão das propriedades estruturais e, até mesmo, das características evolutivas de uma proteína específica.

As lectinas, proteínas amplamente distribuídas na natureza, são encontradas em múltiplas formas moleculares, sendo que grande maioria delas podem ser obtidas em elevadas proporções e com alto grau de pureza pela utilização de diferentes etapas cromatográficas (KENNEDY *et al.*, 1995; LORIS, 2002). Assim, centenas de moléculas desta classe proteica já foram isoladas de vírus, bactérias, fungos, invertebrados, vertebrados e de plantas (NASCIMENTO *et al.*, 2012).

Dentre todas as lectinas isoladas, as obtidas da família botânica *Fabaceae* são uma das mais bem estudadas até o momento (RIBEIRO *et al.*, 2014; SHARON; LIS, 1990). As lectinas de *Fabaceae* são conservadas evolutivamente e formam um grupo de proteínas estruturalmente semelhantes. Estas já foram relatadas em diferentes órgãos vegetais, entretanto, as lectinas dessa família são purificadas com mais facilidade a partir de sementes, que compõem cerca de 10% do conteúdo total de proteínas (LAM; NG, 2011; LORIS *et al.*, 1998).

Algumas das tribos que integram a família *Fabaceae* possuem poucas lectinas estudadas, principalmente em grupos mais primitivos, como por exemplo, a tribo *Sophoreae* e a *Swartzieae*, ambas da subfamília *Papilionoideae*. Atualmente, apenas a SLL (*Swartzia laevicarpa* Lectin), do gênero *Swartzia*, pertencente à subtribo *Swartzieae*, foi purificada. O estudo deste gênero taxonômico pode contribuir para o entendimento das relações evolucionárias das lectinas, provenientes da família *Fabaceae*, no que diz respeito ao processamento proteico e às variações estruturais (FERNANDES *et al.*, 2011).

Uma das espécies que integra o gênero e a subtribo acima mencionados é a *Swartzia grandiflora* Bong. (CARDOSO *et al.*, 2012). Contudo, até o momento, a presença de lectinas nas sementes desta espécie não pôde ser comprovada devido o processo de hemólise dos eritrócitos de coelho, visto quando realizado o ensaio de atividade hemaglutinante em seu

extrato proteico. Tal fato deve-se, provavelmente, devido à grande presença de metabólitos secundários, como isoflavonoides, saponinas triterpênicas e diterpenos aromáticos em suas sementes (MARQUI *et al.*, 2008).

A análise *in silico* das sequências nucleotídicas e de aminoácidos dessas lectinas de difícil purificação pode ser realizada com a ajuda de técnicas de biologia molecular e bioinformática, podendo-se inferir dados de fundamental importância para o seu posterior isolamento, tais como: os sítios de ligação a carboidratos e metais, a presença de N-glicosilações, o pI aproximado da lectina, a presença e a quantidade de pontes dissulfeto, as proteínas de outras espécies que apresentam similaridades evolutivas entre suas sequências, podendo-se, inclusive, realizar uma modelagem tridimensional para a estrutura da lectina a ser purificada.

Neste sentido, o presente estudo foi elaborado, baseando-se no seguinte questionamento: A análise do fragmento nucleotídico que codifica a lectina de *S. grandiflora* possibilitará uma melhor compreensão desta proteína, tais como os sítios de interação a outras moléculas e características evolutivas, fornecendo dados que facilitarão a purificação de sua forma nativa?

Desta forma, diante deste questionamento, formulamos a seguinte hipótese: “A clonagem molecular dos produtos obtidos da RACE, bem como as análises de bioinformática possibilitarão a caracterização *in silico* da sequência codificadora dessa lectina, fornecendo dados sobre sua estrutura, função e contribuindo para o posterior processo de purificação da mesma.”

Diante de todas as informações já expostas, compreendemos que a análise das sequências que codificam a lectina da semente de *Swartzia grandiflora*, aqui denominada SGL, é de considerável importância para o meio acadêmico. Isto se justifica pelo fato de que esta pesquisa poderá fornecer a primeira modelagem molecular feita para uma lectina do gênero *Swartzia*, além de aprimorar o entendimento das relações filogenéticas entre as lectinas de *Fabaceae*, obtendo dados que contribuam para a purificação de sua lectina e, conseqüentemente, o fornecimento de uma molécula de alto valor biotecnológico agregado.

## 2 FAMÍLIA *FABACEAE*

A família *Fabaceae* é constituída por, aproximadamente, 19.500 espécies e 751 gêneros, sendo considerada a terceira maior família, em número de espécies, do grupo das Angiospermas. Adicionalmente, esta família representa um dos maiores exemplos de manipulação e utilização vegetal pela cultura humana, podendo ser utilizada na alimentação humana e de animais domésticos, na utilização da madeira, em ornamentações e obtenção de fármacos de conhecimento tradicional ou para indústrias farmacêuticas (BRUNEAU *et al.*, 2013).

Até o presente momento, a família *Fabaceae* encontra-se dividida em três subfamílias (*Caesalpinioideae*, *Papilionoideae* e *Mimosoideae*), apresentando uma distribuição cosmopolita, que comporta uma rica diversidade de espécies, hábitos e ambientes de ocorrência (BRUNEAU *et al.*, 2013; KOENEN *et al.*, 2013). A utilização de técnicas de biologia molecular vem contribuindo, consideravelmente, para a compreensão da sistemática e da evolução dessas subfamílias (CARDOSO *et al.*, 2012, 2013).

### 2.1 Subfamília *Caesalpinioideae*

A subfamília *Caesalpinioideae* é um grupo parafilético que compõe a base da família *Fabaceae* a qual foi derivada das subfamílias monofiléticas *Mimosoideae* e *Papilionoideae*. As espécies desta diversificada subfamília ocorrem, principalmente, nas regiões tropicais e subtropicais da América do Sul, África e Sudoeste da Ásia. Esta subfamília inclui, aproximadamente, 2250 espécies em 171 gêneros, podendo ser dividida em quatro tribos: *Cercideae*, *Detarieae*, *Cassieae* e *Caesalpinieae* (BRUNEAU *et al.*, 2013).

### 2.2 Subfamília *Papilionoideae*

A espetacular radiação de espécies de *Fabaceae*, da subfamília *Papilionoideae*, que pode ser estimada em 13.800 espécies, é um legado da sua ampla variação ecológica e biogeográfica, associada a uma larga diversificação nas florestas tropicais, savanas, florestas sazonais e regiões temperadas de todo o planeta. As espécies que compõem a subfamília *Papilionoideae* são bem conhecidas pela sua importância econômica para a agricultura e para a indústria alimentícia, como por exemplo, a soja [*Glycine max* (L.) Merr.], feijões (*Phaseolus* L. spp. e *Vicia faba* L.), amendoim (*Arachis hypogaea* L.), lentilha (*Lens culinaris* Medik.),

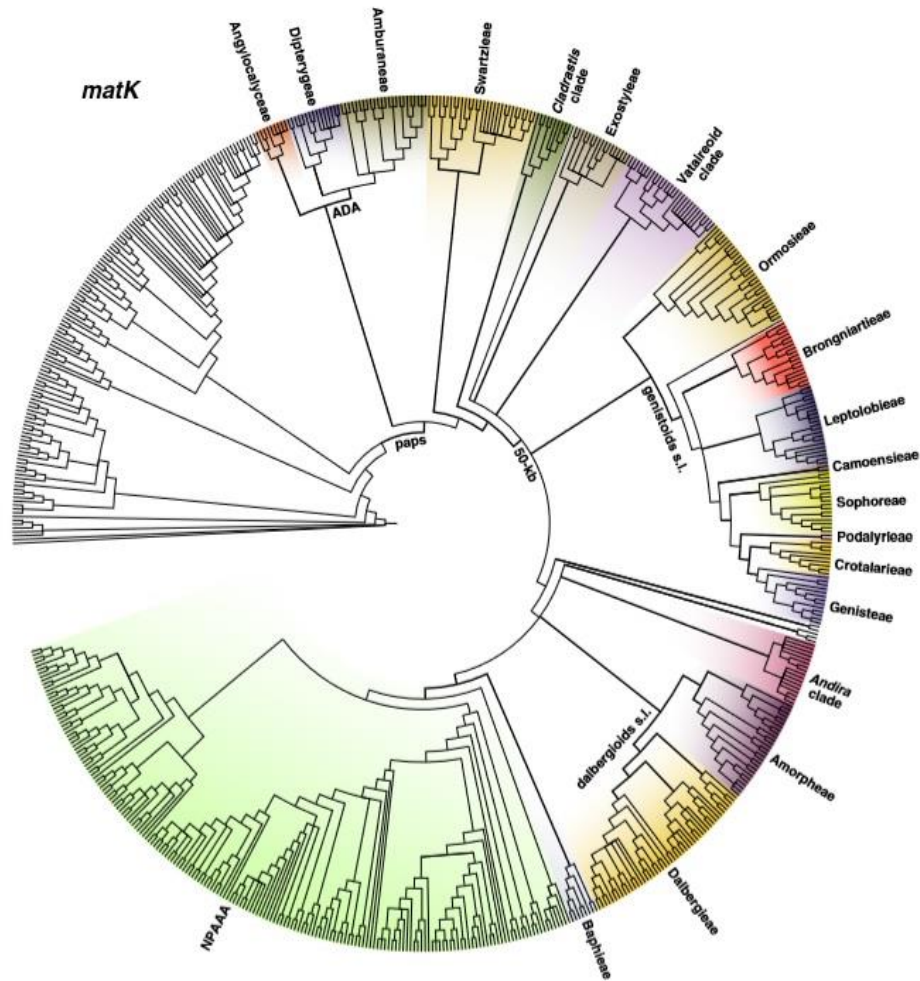


ervilha (*Pisum sativum L.*) e alfalfa (*Medicago sativa L.*), uma importante cultura forrageira (BRUNEAU *et al.*, 2013).

Recentemente, as sequências de 668 genes que codificam a matK de espécies que compõem a subfamília *Papilionoideae* foram realinhadas e uma nova árvore filogenética foi resolvida, com foco na compreensão das relações evolutivas das espécies basais desse grupo taxonômico. A Figura 1 mostra a árvore filogenética das tribos basais da subfamília *Papilionoideae*. Dentre as subtribos que constituem essas espécies basais, encontram-se a *Sophoreae* e a *Swartzieae*, cuja subtribo desta última é formada por oito gêneros, a saber: *Ateleia*, *Bobgunia*, *Bocoa*, *Candolleodendron*, *Cyathostegia*, *Fairchildia*, *Swartzia* e *Trischidium* (CARDOSO *et al.*, 2013)

*Swartzia Schreb* é da tribo *Swartzieae*, o gênero de *Papilionoideae* mais diverso e rico em espécies. Este gênero compreende cerca de 213 espécies amplamente distribuídas nos mais diversos biomas, porém são sobretudo comuns e diversas nas florestas de terras baixas do norte da América do Sul, especialmente na Bacia Amazônica e Escudo das Guianas. No Brasil, o gênero ocorre principalmente na Floresta Amazônica, possuindo, contudo, representantes na Caatinga, Cerrado, Mata Atlântica e Pantanal (TORKE; SCHAAL, 2008).

Figura 1 – Árvore filogenética das tribos basais da subfamília *Papilionoideae*



Fonte: CARDOSO *et al.*, 2013.

### 2.3 Subfamília *Mimosoideae*

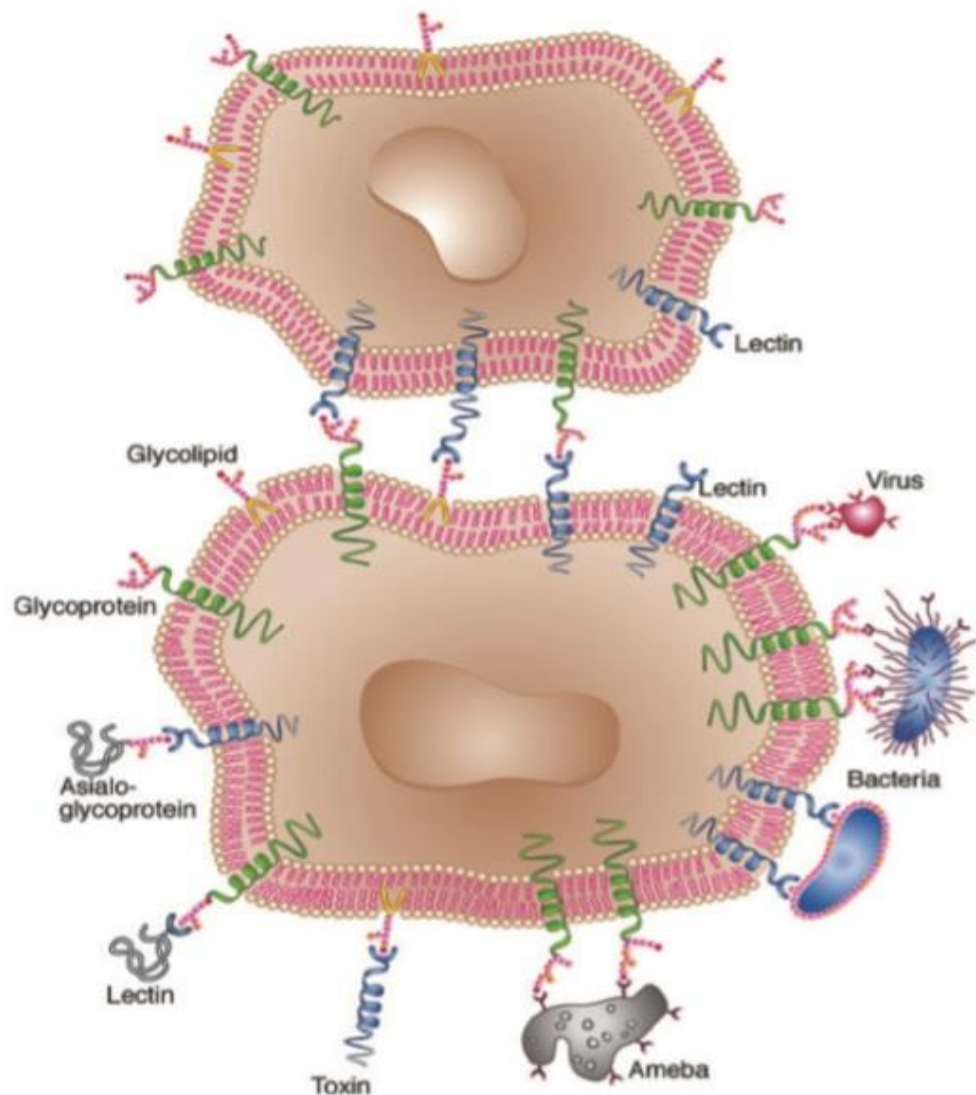
A subfamília *Mimosoideae* possui uma distribuição pantropical e é a segunda maior subfamília de *Fabaceae*, composta por, aproximadamente, 3.271 espécies. Embora a subfamília *Caesalpinioideae* possua mais gêneros que a *Mimosoideae*, o amplo número de espécies desse grupo reflete uma elevada riqueza de espécies na maioria dos seus 83 gêneros, especificamente nos gêneros Acácia e Mimosa (BRUNEAU *et al.*, 2013).

### 3 LECTINAS

Lectinas são proteínas ou glicoproteínas de origem não imunológica que possuem pelo menos um domínio não catalítico que interage de forma reversível e específica a diferentes tipos de glicoconjugados (HE *et al.*, 2015; NASI; PICARIELLO; FERRANTI,

2009). Essa interação entre lectinas e carboidratos ocorre através do Domínio de Reconhecimento a Carboidratos (CRD, sigla proveniente do termo em inglês “Carbohydrate - Recognition Domain”), um domínio representativo desse grupo de proteínas (ROCHA *et al.*, 2015). A propriedade de ligação a mono ou oligossacarídeos confere às lectinas uma grande atividade em processos celulares, como por exemplo, aglutinação, adesão, reconhecimento, sinalização, fertilização, opsonização, mitogênese e apoptose (SINGHA *et al.*, 2015). A Figura 2 mostra os processos de interação entre lectinas e carboidratos da superfície celular (SHARON; LIS, 2004).

Figura 2 – Exemplificação do processo de interação entre lectinas e carboidratos da superfície celular



Fonte: SHARON; LIS, 2004.

### 3.1 Breve Histórico da Lectina

Em 1888, Hermann Stillmark descreveu, pela primeira vez, a atividade de proteínas com propriedades hemaglutinantes em extratos vegetais de *Ricinus communis* e, desde então, diversas destas proteínas foram isoladas e caracterizadas, no que diz respeito a sua estrutura molecular, especificidade por carboidratos e propriedades bioquímicas. As lectinas foram, inicialmente, designadas como “hemaglutininas” ou, mais comumente, “fitohemaglutininas”, visto que são detectadas pela aglutinação de eritrócitos, além de serem, majoritariamente, encontradas em plantas. A capacidade das aglutininas vegetais de distinguir entre eritrócitos de diferentes grupos sanguíneos fez Boyd e Elizabeth Shapley propor, em 1954, o termo lectina, que advém do termo, em Latim, *legere*. Landsteiner & Raubitschek (1908) discutiram, pela primeira vez, a especificidade das lectinas e observaram que vários extratos de sementes de *Fabaceae* apresentavam diferentes propriedades de hemaglutinação, quando analisados com eritrócitos de diferentes espécies animais, provavelmente, porque as lectinas de diferentes espécies vegetais, frequentemente, diferem na sua estrutura molecular e na sua especificidade. A descoberta da especificidade das lectinas por açúcares foi atribuída a James B. Sumner, que isolou, por precipitação salina, e cristalizou uma lectina de sementes de *Canavalia ensiformis* (Con A), sendo esta a primeira até o presente momento (KENNEDY *et al.*, 1995; SHARON; LIS, 2004).

### 3.2 Lectinas de *Fabaceae*

As lectinas são amplamente distribuídas na natureza. Estudos já citaram sua presença em vírus, bactérias, fungos, plantas e animais (FERNANDES *et al.*, 2015; LI *et al.*, 2012). As sementes da família botânica *Fabaceae* são uma das principais fontes de busca deste grupo de proteínas, chegando a compor até 10% do conteúdo total de proteínas. A família das lectinas de *Fabaceae* constituem, aproximadamente, 60% das 361 estruturas depositadas no banco Glyco3D. Este grupo, em particular das lectinas, na maioria das vezes, consiste de duas ou quatro subunidades (25-30 kDa), cada qual com um CRD. A interação entre o CRD e o carboidrato requer a presença de  $\text{Ca}^{2+}$  e  $\text{Mn}^{2+}$  (ou outro metal de transição). A estrutura primária das lectinas de sementes de *Fabaceae* apresenta uma notável homologia, demonstrando que estas proteínas foram, evolutivamente, conservadas. A estrutura tridimensional dessa família de lectinas é similar e caracterizada pelo elevado conteúdo de folhas  $\beta$  em relação à presença de  $\alpha$  hélices (CARLINI; GROSSI-DE-SÁ, 2002; CARVALHO

*et al.*, 2015; NASCIMENTO *et al.*, 2012; RIBEIRO *et al.*, 2014).

### 3.3 Classificação das Lectinas de *Fabaceae*

A elucidação da estrutura tridimensional de grande parte das lectinas da família *Fabaceae* contribui para a classificação deste grupo proteico quanto ao número de domínios de ligações a carboidratos, sendo estas agrupadas da seguinte maneira:

- a) merolectinas: apresentam um único domínio de ligação a carboidrato e são incapazes de promover aglutinação celular;
- b) hololectinas: apresentam dois domínios de ligação a carboidratos, idênticos ou homólogos, e são capazes de promover aglutinação celular;
- c) quimerolectinas: são constituídas por um ou mais domínios de ligação a carboidratos e um domínio de função distinta;
- d) superlectinas: são constituídas de, no mínimo, dois domínios de ligação a carboidratos, porém, cada domínio reconhece carboidratos estruturalmente diferentes.

As lectinas de sementes de *Fabaceae* também podem ser classificadas de acordo com a sua especificidade de ligação a carboidratos, compreendendo aquelas ligantes de manose, manose/glicose, manose/maltose, galactose/N-acetilgalactosamina, fucose e ácido siálico (KENNEDY *et al.*, 1995; LORIS, 2002; PEUMANS; VAN DAMME, 1995).

### 3.4 Função

Lectinas de plantas são consideradas proteínas de reserva que podem apresentar propriedades biológicas associadas à proteção do tecido vegetal contra pestes e patógenos, como também a estresses abióticos (DE SOUZA CANDIDO *et al.*, 2011). Tal característica ocorre, principalmente, devido a sua propriedade de interação com glicoconjugados ou a superfície de células patogênicas em solução. Atualmente, é sugerido que as lectinas adquirem sua função de estoque como função secundária, com a atividade de defesa, começando a se tornar a característica principal para este grupo de proteínas. Adicionalmente, as lectinas vegetais têm sido caracterizadas por possuírem uma importante função sobre fungos, bactérias e insetos (PEUMANS; VAN DAMME, 1995).

A membrana celular destes organismos, normalmente, apresenta glicoconjugados de diferentes naturezas, como por exemplo, as glicoproteínas, os glicolipídeos e os

polissacarídeos, que podem possuir uma estrutura similar aos carboidratos, os quais interagem com as lectinas. Desta forma, pode-se inferir que as lectinas são capazes de atuarem como proteínas de defesa, por meio da interação aos glicoconjugados destes microrganismos. Além disso, as lectinas de plantas também possuem a característica de serem estáveis às variações de pH e temperatura, bem como apresentarem resistência à proteases de animais e de insetos. Estas características fazem das lectinas de leguminosas uma potencial fonte de inibição de microrganismos patogênicos (DE SOUZA CANDIDO *et al.*, 2011).

As possíveis aplicações biotecnológicas para as funções e características já citadas das lectinas são diversas e de grande importância, tais como os efeitos antitumoral e antiviral, além da detecção, isolamento e estudo de glicoconjugados (LAM; NG, 2011; UNITT; HORNIGOLD, 2011). Outras aplicações para o grupo proteico em questão encontram-se descritas na Tabela 1.

Tabela 1 – Aplicações biotecnológicas relatadas na literatura científica para algumas lectinas

<b>Lectina</b>	<b>Aplicação biotecnológica</b>	<b>Referências científicas</b>
ASA I e ASA II	Atividade inseticida contra <i>Acyrtosiphon pisum</i>	(FITCHES <i>et al.</i> , 2008)
MuBL	Atividade inseticida contra <i>Aedes aegypti</i>	(SÁ <i>et al.</i> , 2009)
TEL	Inibidora do fungo <i>Microsporium canis</i>	(PINHEIRO <i>et al.</i> , 2009)
ZmCORp	Inibidora do fungo <i>Aspergillus flavus</i>	(BAKER <i>et al.</i> , 2009)

Fonte: elaborada pelo autor.

### 3.5 Características moleculares

A presença de isolectinas em uma única planta já foi relatada na literatura científica. Esta característica deve-se, provavelmente, ao fato das lectinas de leguminosas comporem uma família multigênica, podendo assim ser expressas em diferentes tecidos ou estágios de desenvolvimento. Outra justificativa para a presença de múltiplas formas de uma mesma lectina pode ser atribuída às modificações de cadeias durante o isolamento da proteína, como por exemplo, a hidrólise do grupamento amida da glutamina ou asparagina. Desta maneira, o termo isoforma é o mais adequado, quando distintas formas moleculares de uma lectina ocorrem na mesma espécie (KENNEDY *et al.*, 1995).

As lectinas de *Fabaceae* são, geralmente, sintetizadas como um precursor, que subsequentemente (após a remoção inicial do peptídeo sinal da região N-terminal) sofrem um processamento pós-traducional para, então, transformarem-se na proteína madura. Este processamento pode consistir na clivagem proteolítica da cadeia precursora, corte da região C - terminal, remoção de carboidratos covalentes e, até mesmo, a ligação da região N - e C - terminal original (LORIS *et al.*, 1998).

### 3.6 Clonagem molecular

O estudo das duas lectinas de *Robinia pseudoacacia* (RPbAI e RPbAII) é um exemplo que pode ser citado a fim de exemplificarmos a importância destas análises moleculares das lectinas (VAN DAMME *et al.*, 1995). Três classes de clones de cDNA correspondentes aos polipeptídios a (31.5 kDa) e b (29 kDa) da RPbAI e o polipeptídio c (26 kDa) da RPbAII foram identificadas. O clone LECRPA1 codifica uma sequência com 285 aminoácidos, contendo uma região idêntica ao N - terminal do polipeptídio a. O *pI* estimado para o polipeptídio codificado pela LECRPA1 é 5.04. A sequência de aminoácidos que codifica a LECRPA1 possui dois possíveis sítios de glicosilação nas posições 147 e 188. A LECRPA2 codifica uma sequência com 286 aminoácidos com uma região idêntica ao N - terminal do polipeptídio b, possui *pI* estimado em 4.95 e um possível sítio de glicosilação que coincide com o primeiro possível sítio de glicosilação da LECRPA1. O terceiro clone de cDNA, LECRPA3, codifica uma sequência contendo 272 aminoácidos com um possível sítio de iniciação na posição 13. Apesar da sequência N - terminal determinada para o polipeptídio c e a sequência deduzida para o clone de cDNA da LECRPA3 não serem idênticas, acredita-se que a LECRPA3 codifica um polipeptídio de lectina, já que é visto uma elevada homologia com as sequências da LECRPA1 e LECRPA2 (RABIJNS *et al.*, 2000).

### 3.7 Lectinas da tribo Swartzieae

A primeira descrição para uma lectina desta tribo foi relatada em sementes de *Swartzia pickelli* (CAVALCANTI; COELHO, 1990), nas quais foram descritas a presença de atividade hemaglutinante no extrato destas sementes quando testadas com sangue de diferentes animais. Já no estudo desenvolvido por FERNANDES *et al.*, 2011, foi demonstrado evidências da presença de uma lectina em extratos de sementes de *Swartzia laevicarpa*, após cromatografia de troca iônica. Tal lectina foi purificada por cromatografia de afinidade em

matriz, contendo lactose imobilizada. A SLL (*Swartzia laevicarpa Lectin*) apresentou afinidade à lactose, à galactose e ao N-acetil-D-galactosamina. Além disso, a mesma demonstrou, também, estabilidade à elevadas temperaturas e uma massa molecular de, aproximadamente, 29 kDa. Adicionalmente, foi comprovado que essa lectina é uma glicoproteína. Isto porque quando submetida ao sequenciamento da região N-terminal por degradação de Edman, a SLL apresentou similaridade com as lectinas pertencentes às tribos primitivas da subfamília *Papilionoideae*, a saber, a *Dalbergieae* e a *Sophoreae* (FERNANDES *et al.*, 2012).

## 4 OBJETIVOS

A presente dissertação de Mestrado foi desenvolvida com o intuito de serem alcançados os objetivos das subseções abaixo descritas.

### 4.1 Objetivo geral

Caracterizar, por análises de biologia molecular e bioinformática, as sequências de cDNA, obtidas a partir das sementes de *Swartzia grandiflora* que codificam lectinas.

### 4.2 Objetivos específicos

- Amplificar as regiões que codificam as Unidades de Término de Replicação da porção (UTR) 3'e 5' de lectinas do RNA total das sementes de *Swartzia grandiflora*;
- Clonar os fragmentos descritos acima em devidos vetores de clonagem, bem como sequenciar suas respectivas bases nucleotídicas;
- Analisar as características das sequências obtidas por meio de ferramentas de bioinformática, gerando um modelo tridimensional para a SGL, a partir dos dados obtidos.

## 5 MATERIAL

O material utilizado no decorrer desta pesquisa científica encontra-se descrito nas subseções abaixo.



### 5.1 Sementes de *Swartzia grandiflora*

As sementes foram coletadas no primeiro semestre de 2014, advindas de uma árvore que se encontra localizada no Campus do Pici, da Universidade Federal do Ceará, sendo reconhecida pelo Herbário Prisco Bezerra (número de voucher: 54159). A flor, a vagem, e a semente de *Swartzia grandiflora* estão representadas na Figura 3.

Figura 3 – Ramo de *Swartzia grandiflora* com botão floral em desenvolvimento (A); ramo de *Swartzia grandiflora* com flor desenvolvida (B) e semente de *Swartzia grandiflora*



Fonte: José Edvar (2014).

### 5.2 Microrganismos

A clonagem da lectina de *Swartzia grandiflora* foi realizada utilizando-se como hospedeiro células de *Escherichia coli* (*E. coli*) TOP10F' (Invitrogen).

### 5.3 Plasmídeos

O pGEM – T (Promega) foi utilizado como vetor em células de *E. coli* para a clonagem dos fragmentos de interesse desta pesquisa. O mapa do vetor de clonagem encontra-se representado no ANEXO A.

### 5.4 Enzimas

As enzimas RQ1 RNase-FreeDNase 1 U/μL da Promega, Taq DNA polimerase 5 U/μL da GE HealthCare, Improm – II Reverse Transcriptase 5 U/μL da Promega, a T4 DNA ligase 3 U/μL da Promega, a PvuII 10 U/μL da ThermoScientific, e as enzimas CIP (Fosfatase Alcalina do Intestino do Bezerro), TAP (Pirofosfatase Ácida do Tabaco), T4 RNA ligase e a M-MLV RT (MoloneyMurineLeukemiaVirus Reverse Transcriptase) foram provenientes do Kit FirstChoice RLM-RACE da Ambion.

### 5.5 Oligonucleotídeos

Os oligonucleotídeos usados como iniciadores nas reações de PCR-RACE foram sintetizados pela Eurofins MWG/Operon, a saber: SGLForward (5' - CAR GAR GAY YTN YTN YTN CAR GGN GAY GC-3'), SGLReverse (5' - GAAACCAACTCTAACCCACTC NGG - 3'), o 3' RACE Outer Primer (5'- GCGAGCACAGAATTAATACGACT-3') e 3' RACE Adapter (5'-GCGAGCACAGAATTAATACGACTCACTATAGGT12VN-3'). Além destes, foram utilizados os oligonucleotídeos presentes no Kit FirstChoice RLM-RACE (Ambion), sendo eles: 5' RACE Outer Primer (5'-GCTGATGGCGATGAATGAACACTG-3'), o 5' RACE Adapter (5'- GCUGAUGGCGAUGAAUGAACACUGCGUUUGCUGGCUUUGAUGAAA-3') e os RandomDecamers.

### 5.6 Kits e reagentes

Os meios de cultura sólido e líquido para crescimento de *E. coli* foram preparados segundo SAMBROOK; FRITSCH; MANIATIS, 1989. Os kits ImProm-II™ Reverse Transcription System e NucleoSpinPlasmid (NoLid) foram adquiridos da Promega e Macherey-Nagel, respectivamente. O marcador molecularScada DNA Ladder 1 kb foi proveniente da Ludwig Biotec. O kit FirstChoice® RLM-RACE e o inibidor de RNase foram

provenientes da Ambion. Antibióticos (estreptomicina e carbenicilina), IPTG (isopropil -  $\beta$  - D - tiogalactopiranosídeo), X - GAL (5 - bromo - 4 - chloro - 3 - indolyl - beta - D - galacto - pyranoside) e demais reagentes foram todos de grau analítico.

## 6 MÉTODOS

As metodologias utilizadas no decorrer desta pesquisa científica encontram-se descritas nas subseções abaixo.

### 6.1 Desenho e obtenção dos iniciadores

A região N-terminal da SLL, constituída pelos seguintes aminoácidos: 1ADIVSFTYTKFRPNQEDLLLQGDAEVVSSGKLQLTAVDD39, foi utilizada como modelo para o desenho dos iniciadores SgLFoward e SgLReverse, por ser a única lectina do gênero *Swartzia* com a região N-terminal obtida (FERNANDES *et al.*, 2012). Neste sentido, a sequência de aminoácidos da região N-terminal dessa lectina foi submetida ao programa tblastn (<http://blast.ncbi.nlm.nih.gov>) do NCBI, o qual faz uma busca por nucleotídeos, utilizando uma sequência proteica como fonte de busca. Os genes codificadores de Sophorajaponica clone LECSJAsglectin precursor mRNA (U63011), Sophorajaponica clone LECSJAbmIlectin precursor mRNA (U63012), Sophorajaponica clone LECSJAbmIlectin precursor mRNA (U63013), Sophorajaponica clone LECSJAbglectin precursor mRNA (U63014), SophorajaponicalectinproteinmRNA (KC140286), SophoraflavescenslectinmRNA (AF285121), SophoraalopecuroideslectinmRNA (DQ011517) e CladrastisluteastorageproteinmRNA (U21940) foram escolhidos dentre todas as sequências que apresentaram similaridade à região N-terminal da SLL e por serem de um grupo taxonômico filogeneticamente próximo ao gênero *Swartzia*. Posteriormente, estas sequências foram convertidas para aminoácidos pelo programa Translate Tools ([www.expasy.org](http://www.expasy.org)), que indica qual dos frames codifica uma proteína funcional. As regiões nucleotídicas, aproximadamente, correspondentes ao início e ao término de cada uma das sequências obtidas foram pareadas manualmente. Os iniciadores foram, então, montados, levando em consideração o grau de conservação e degeneração de cada um dos nucleotídeos que compõem o início e o fim da proteína madura. A Figura 4 mostra os alinhamentos para elaboração dos iniciadores utilizados.

Figura 4 – Representação do desenho dos iniciadores SGLFoward e SGLReverse

Alinhamento para elaboração do iniciador SgLFoward	
	<b>Q E D L L L Q G D A</b>
U63011:	CAAGAGGATCTACTCCTCCAAGGTGATGCC <b>Q E D L L L Q G D A</b>
U63014:	CAAGAGGATCTACTCCTCCAAGGTGATGCC <b>Q E D L L L Q G D A</b>
AF285121:	GAGGATCTGTTATTCAGGGTGATGCA <b>Q N E E D L L F Q G D A</b>
DQ011517:	CAAAATGAAGAGGATCTGTTATTCAGGGTGATGCC <b>Q F N F E D L I L Q R D A</b>
U63013:	CAGCCAAACCAGAAGATCTCATCTCCAACGTGATGCC <b>Q D E L L L Q G D A</b>
U21940:	CAAGACGAATCTACTCCTCCAAGGCGATGCC <b>Q R D L I L Q G D A</b>
U63012:	CAAAGGGATCTAATCCTCCAAGGTGATGCC <b>Q F Y L I L Q G D A</b>
KC140286:	CAACCATATCTGATCTTGCAAGGTGATGCC
SgLFoward:	CARGARGAYTNYHTNCARGGNGAYGC
Alinhamento para elaboração do iniciador SgLReverse	
	<b>F E W V R V G F</b>
U63011:	CCGGAGTGGGTTAGATTGGTTTC <b>F E W V R V G F</b>
U63014:	CCAGAGTGGGTTAGATTGGTTTC <b>F E F V R V G F</b>
AF285121:	CCGGAATTTGTCAGAGTTGGATTC <b>F E W V R V G F</b>
DQ011517:	CCGGAATGGGTCAGGTTGGATTC <b>F E W V R V G F</b>
U63013:	CCTGAATGGGTTAGATTGGATTC <b>F E W V R V G F</b>
U21940:	CCGGAGTGGGTTAGATTGGTTTC <b>F E W V R I G F</b>
U63012:	CCAGAATGGGTTAGAATTGGATTC <b>F E W V R F G F</b>
KC140286:	CCGGAGTGGGTAAGATTGGTTTC
SgLReverse:	CCNGARTGGGTNAGNGTNGNNTTY

Fonte: elaborada pelo autor.

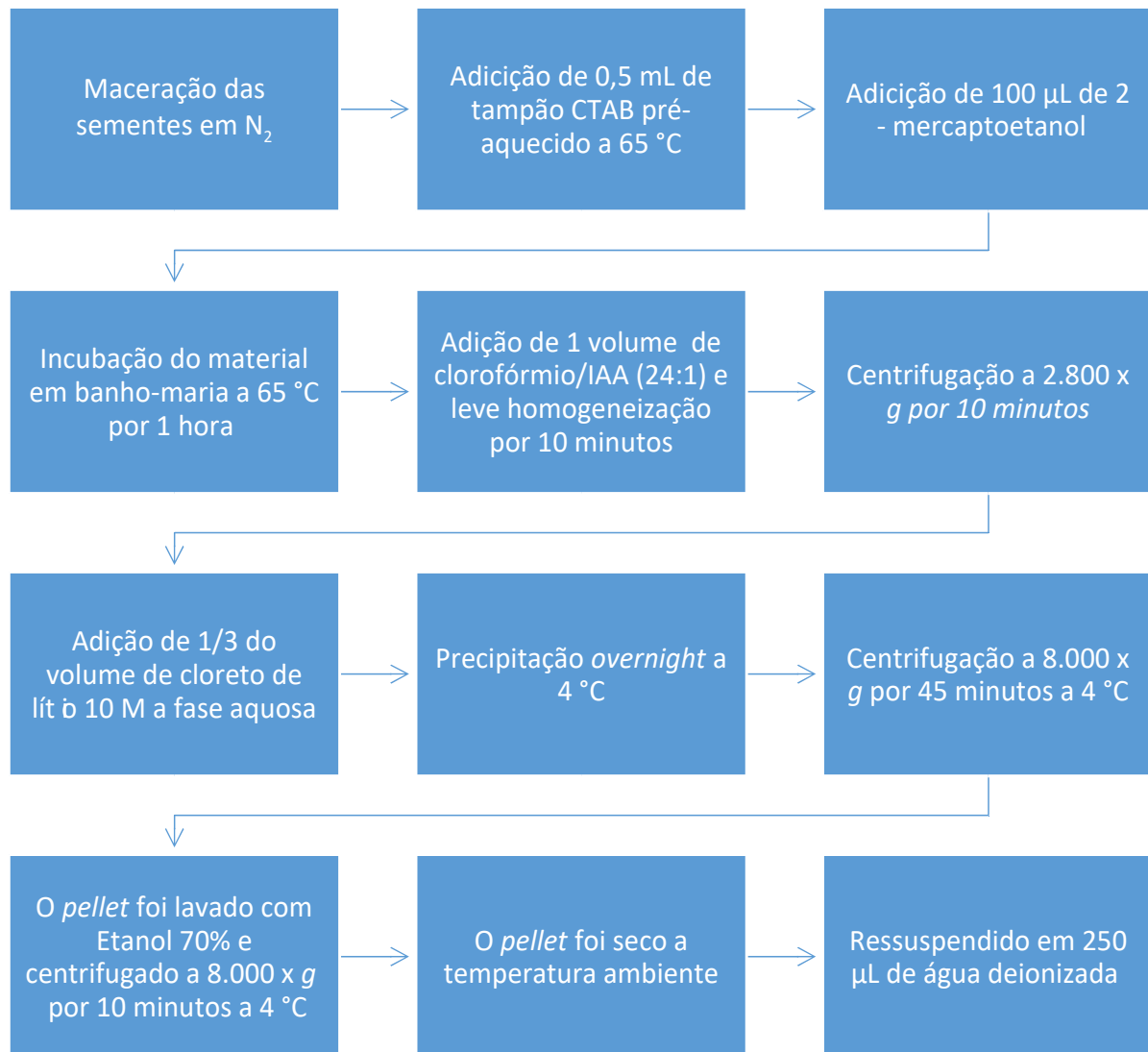
## 6.2 Maceração das sementes de *Swartzia grandiflora*

Sementes de *Swartzia grandiflora* foram coletadas de suas vagens e, imediatamente, submetidas a um processo de maceração em almofariz com pistilo, utilizando-se nitrogênio líquido. Após este processo, obtivemos um material finamente pulverizado, o qual foi armazenado em frações de 1g cada, em tubos de 15 mL, que foram conservados em ultrafreezer a -80°C, até o momento da extração do RNA total.

## 6.3 Extração de RNA total

O RNA total das sementes de *Swartzia grandiflora* foi extraído com base na metodologia descrita por CHANG; PURYEAR; CAIRNEY, 1993, com pequenas modificações. A Figura 5 demonstra o passo a passo da referida metodologia.

Figura 5 – Representação esquemática da extração de RNA das sementes de *Swartzia grandiflora*, utilizando a metodologia de precipitação por Cloreto de Lítio 10M (CHANG; PURYEAR; CAIRNEY, 1993)



Fonte: Paulo Abraão (2014).

Como vimos acima, o tampão CTAB foi pré-aquecido a  $65^\circ C$  e, em seguida, este foi adicionado a 1g de amostra previamente pulverizada, na proporção de 0,5mL de tampão para cada 0,1g de semente pulverizada. Logo após, foram adicionados 100 $\mu L$  de 2-mercaptoetanol e a amostra foi incubada em banho-maria a  $65^\circ C$ , por 1 hora, sob leve agitação, a cada 10 minutos. Passado este período, foi adicionado 1 volume (cerca de 5mL) de clorofórmio/IAA (24:1) e, em seguida, a amostra foi, suavemente, homogeneizada por 10 minutos, à temperatura ambiente e, posteriormente, centrifugada a 2.800 x g por 10 minutos. O sobrenadante obtido do passo anterior foi transferido para um novo recipiente contendo

5mL de clorofórmio/IAA (24:1), agitado por mais 5 minutos e centrifugado a 2.800 x g por 10 minutos. A fase aquosa superior foi transferida para um novo tubo, ao qual foram adicionados, aproximadamente, 1/3 de volume de cloreto de lítio, a partir de um estoque a 10M. Este material foi deixado em repouso (*overnight*) a 4°C, para precipitação. No dia seguinte, o material foi submetido à centrifugação a 4°C, por 45 minutos a 8.000 x g. O sobrenadante foi descartado e o precipitado (pellet) foi lavado com etanol 70%, por meio de centrifugação a 8.000 x g por 10 minutos a 4°C. Posteriormente, o pellet foi seco à temperatura ambiente por, aproximadamente, 30 minutos e ressuspenso em 250 µL de água livre de RNases. O RNA total foi, então, quantificado por espectrofotometria no aparelho GeneQuantda AmershamBiosciences, sendo considerada uma unidade de DO260nm equivalente a 40 µg/mL de RNA. O grau de pureza foi determinado com base na relação de absorbância A260nm/A280nm. Para esta análise, foi diluído 2µL da amostra do RNA total em 98µL de água deionizada. Logo após, a amostra foi armazenada a -80°C.

#### **6.4 Análise da integridade do RNA total**

O RNA total extraído das sementes de *Swartzia grandiflora* teve sua integridade avaliada por eletroforese. Inicialmente, antes da eletroforese, foi preparada a reação para um volume de 10µL, adicionando-se 60% (v/v) de azul de bromofenol [0,25% (m/v), glicerol 30% (v/v), preparados com tampão TE (Tris-EDTA)], 1µL de brometo de etídio (1mg/mL) e 3µL da amostra de RNA (aproximadamente 1µg). Em seguida, a amostra foi aplicada em gel de agarose 1,5% preparado com tampão TAE (Tris-acetato 0,04 M, EDTA 1 mM) 10% e submetidas a uma corrida eletroforética a 63 volts. A presença de bandas foi visualizada por meio de um transiluminador (VilberLourmat) de UV e uma câmera acoplada ao microcomputador dotado de *software* de captura e análise de imagens.

#### **6.5 Tratamento do RNA com a DNase**

Amostras de 10µg de RNA total foram padronizadas para um volume final de 30µL com água livre de nuclease, no intuito de diminuir a presença de DNA genômico. Cada alíquota de RNA foi tratada com 10U da enzima DNase (RQ1 DNase, Promega) e 3µL de tampão da enzima 10 X, a 37°C, durante 30 minutos. A inativação da DNase foi alcançada após a adição de 3µL de RQ1 DNase Stop Solution e, posteriormente, uma incubação a 65°C por 10 minutos foi realizada. O RNA total tratado foi, em seguida, estocado a -20°C.

## **6.6 Amplificação rápida das extremidades do cDNA (RACE)**

Os procedimentos descritos nas subseções, a seguir, descrevem como foram realizadas as amplificações das regiões 5' UTR e 3' UTR da SgL.

### **6.6.1 Amplificação da região 5' UTR**

Os procedimentos descritos nas subseções, a seguir, descrevem como foi realizada a amplificação da região 5' UTR da SgL.

#### **6.6.1.1 Tratamento do RNA total com a enzima CIP**

Inicialmente, ao RNA total purificado (10µg) foram adicionados 2µL de tampão CIP 10 X, 2µL da enzima CIP e água livre de RNase para um volume final de 20µL. A mistura foi homogeneizada, suavemente, e incubada a 37°C, durante 1 hora. Em seguida, foram acrescentados à mistura 15µL de solução de acetato de amônio, 150µL de fenol:clorofórmio (1:1) e 115µL de água livre de RNase. Posteriormente, a mistura foi homogeneizada e centrifugada a 10.000 x g, à temperatura ambiente, durante 5 minutos. A fase aquosa foi transferida para um microtubo de 1,5mL, livre de RNase, e adicionou-se 150µL de clorofórmio P.A. Depois, a mistura foi homogeneizada e centrifugada nas mesmas condições anteriores. O sobrenadante foi descartado e ao pellet foram adicionados 150µL de isopropanol P.A. Após agitação em vortex, a mistura foi incubada em banho de gelo, durante 10 minutos. Depois de incubada, a amostra foi, imediatamente, centrifugada a 13.500 x g, por 20 minutos, à temperatura ambiente. Após o descarte do sobrenadante, o pellet formado foi lavado com 500µL de uma solução de etanol 70% e centrifugado na mesma velocidade anterior, por 5 minutos. O sobrenadante foi descartado. Finalmente, o pellet de RNA tratado com a enzima CIP foi eluído em 11µL de água livre de RNase e estocado a -20°C, até o momento do tratamento com a enzima TAP.

#### **6.6.1.2 Tratamento do RNA total com TAP**

Após o tratamento do RNA total com a enzima CIP, sucedeu-se ao tratamento com a enzima TAP, de acordo com o protocolo descrito a seguir. Inicialmente, a 5µL do RNA total tratado com CIP foram adicionados 1µL de tampão TAP 10 X, 2µL da enzima TAP e água

livre de RNase para um volume final de 10 $\mu$ L. A mistura foi homogeneizada, suavemente, e incubada a 37°C, durante 1 hora. Em seguida, foi estocada a -20°C, até a realização da reação de ligação do adaptador (oligonucleotídeo 5' RACE Adapter) ao RNA tratado com as enzimas CIP e TAP.

#### **6.6.1.3 Reação de ligação do RNA tratado ao iniciador 5' RACE Adapter**

Previamente, a 2 $\mu$ L de RNA total tratado com as enzimas CIP e TAP, foram adicionados 1 $\mu$ L do iniciador 5' RACE Adapter (0,3 $\mu$ g), 1 $\mu$ L de tampão RNA ligase 10 X, 2 $\mu$ L da enzima T4 RNA ligase e água livre de RNase para um volume final de 10 $\mu$ L. A mistura foi homogeneizada, suavemente, e incubada a 37°C, durante 1 hora. Em seguida, foi estocada a -20°C, até o momento da reação com a transcriptase reversa (RT-PCR).

#### **6.6.1.4 Transcrição reversa do material ligado ao iniciador 5' RACE Adapter**

O cDNA foi sintetizado por meio de uma reação de polimerização, catalisada pela M-MLV RT, conforme protocolo descrito a seguir. Inicialmente, a 2 $\mu$ L de RNA ligado ao adaptador 5' RACE foram adicionados 4 $\mu$ L de dNTPs 10mM, 2 $\mu$ L de RandomDecamers (primers constituídos de 10pb), 2 $\mu$ L de tampão da transcriptase reversa 10 X, 2 $\mu$ L de inibidor de RNase, 1 $\mu$ L de enzima M-MLV RT e água livre de RNase para um volume final de 20 $\mu$ L. A mistura foi homogeneizada, suavemente, e incubada a 50°C, durante 1 hora. Em seguida, uma alíquota de cDNA sintetizado foi analisado através de eletroforese em gel de agarose a 1,2% em tampão TAE, nas mesmas condições do item 4.4, seguida da visualização por meio de um transiluminador de UV e uma câmera acoplada ao microcomputador, dotado de software de captura e análise de imagens. Os cDNAs foram estocados a -20°C, até o momento da reação de Nested PCR para 5' RLM-RACE.

#### **6.6.1.5 Reação Nested PCR para obtenção da região 5' UTR**

A 1 $\mu$ g do cDNA anteriormente obtido, foram adicionados 2,5 $\mu$ L de tampão 10 X da enzima Taq DNA polimerase (GE Healthcare), 0,5 $\mu$ L de dNTP 10mM, 1,25 $\mu$ L do iniciador forward 5' RACE Outer primer 10 $\mu$ M, 5 $\mu$ L do iniciador SGLReverse 5 $\mu$ M, 1,5 $\mu$ L de cloreto de magnésio 25mM, 0,4 $\mu$ L de Taq DNA polimerase (2U) e água livre de RNase para um volume final de 25 $\mu$ L (Tabela 2). A amplificação foi realizada nas mesmas condições



apresentadas na Tabela 3. Alíquotas dos produtos obtidos foram corados com azul de bromofenol, assim como no item 4.4, e analisados através de eletroforese em gel de agarose 1,2% e visualizados com brometo de etídio. A captura da imagem foi realizada com o auxílio de luz UV (302nm) em um transiluminador (Macrovue™ Transluminator, Pharmacia Biotec).

Tabela 2 – Componentes utilizados na amplificação da região 5' UTR da SgL, por Nested PCR

Reagentes	Volumes (µL)	Concentração final
Tampão 10 X	2,5	1 X
dNTPs 10 mM	0,5	200 µM
5'RACE Outer Primer 10 µM	1,25	0,5 µM
SgLReverse 5 µM	2,5	0,5 µM

Fonte: elaborada pelo autor.

Tabela 3 – Condições utilizadas para a amplificação das regiões 5' UTR e 3' UTR SgL, por PCR

	Estágio	Repetição	Temperatura (°C)	Tempo
<b>Desnaturação inicial</b>	1	1	95	3 minutos
<b>Amplificação</b>	2	32	95	45 segundos
			49	45 segundos
			72	1,5 minutos
<b>Extensão final</b>	3	1	72	8 minutos

Fonte: elaborada pelo autor.

## 6.6.2 Amplificação da região 3' UTR

Os procedimentos descritos nas subseções, a seguir, descrevem como foi realizada a amplificação da região 3' UTR da SgL.

### 6.6.2.1 Transcrição reversa do RNA de *S. grandiflora* para a obtenção da região 3' UTR

A obtenção do cDNA utilizado para a amplificação da região 3' UTR foi realizada combinando, inicialmente, 3µL de RNA de *S. grandiflora* (~800ng) com 2µL de 3' RACE

Adapter (0,5µg) e 6,6µL de água livre de RNase. Esse *mix* reacional foi submetido a 70°C por 5 minutos e, então, a 4°C por 5 minutos. Alíquotas de 4µL de tampão Improm-II 5X, 2,4µL de MgCl<sub>2</sub> 25mM, 1µL de dNTP 10mM e 1µL da transcriptase reversa da Improm-II foram acrescentados ao microtubo, contendo o *mix* reacional. Logo após, este *mix* foi submetido a determinado período de tempo em termociclador, onde realizou-se o anelamento a 25°C por 5 minutos, extensão a 42°C por 60 minutos e inativação da transcriptase reversa a 70°C por 15 minutos.

### 6.6.2.2 Race 3'

A amplificação da região 3' UTR da SgL foi, então, realizada por PCR convencional. Uma quantidade de 980ng de cDNA, obtido do procedimento descrito anteriormente, foi adicionada a um microtubo estéril contendo 2,5µL de tampão da enzima Taq DNA polimerase (GE Healthcare), 1,5µL de cloreto de magnésio 25mM, 0,5µL de dNTP 10mM, 2,5µL de primer forward 5µM (SgLForward), 1,25µL de primer reverse 10µM (3'RACE Outer primer), 0,4µL de Taq DNA polimerase (2U) e água (q.s.p.) livre de RNase, para um volume final de 25µL (Tabela 4). As condições de amplificação foram as mesmas da Tabela 3. O produto de PCR obtido foi submetido à eletroforese em gel de agarose 0,8% a fim de averiguar a eficiência da amplificação realizada. Para tanto, uma alíquota desta amostra foi corada com azul de bromofenol, assim como no item 4.4. A presença de banda foi avaliada por meio de eletroforese em gel de agarose 0,8% e visualizada com ajuda de um transiluminador de UV após a corrida eletroforética.

Tabela 4 – Componentes utilizados na amplificação da região 3' UTR da SgL, por PCR

Reagentes	Volumes (µL)	Concentração final
Tampão 10 X	2,5	1 X
dNTPs 10 mM	0,5	200 µM
SgLForward 5 µM	2,5	0,5 µM
3'RACE Outer Primer 10 µM	1,25	0,5 µM
cDNA	2,0	980 ng
MgCl <sub>2</sub> 25 mM	1,5	1,5 mM

Fonte: elaborada pelo autor.

## 6.7 Clonagem

As metodologias utilizadas para a clonagem encontram-se descritas nas subseções abaixo.

## 6.8 Ligação

A ligação do produto obtido foi realizada em vetor plasmidial, utilizando o sistema pGEM-T. Inicialmente, um spin foi realizado no tubo contendo o plasmídeo e, então, 2µL (100ng) do pGEM-T foram adicionados a 10µL de tampão de ligação 2 X, 2µL das amostras obtidas da PCR-RACE (~1µg) e 2µL da enzima T4 DNA ligase (6U). Logo após, o *mix* da ligação foi incubado por 16 horas a 4°C. O produto da ligação foi utilizado para transformação das células de *E. coli* TOP10F'.

## 6.9 Transformação

As células de *E. coli* da cepa TOP10F' foram, primeiramente, tornadas competentes através do método de eletroporação. Para tanto, 100µL de uma cultura estoque mantida em glicerol, a -80°C, foram inoculados em uma placa de petri, contendo 25mL de LB ágar (peptona de caseína 1%, extrato de levedura 0,5%, NaCl 1%, pH 7,2), contendo estreptomicina (50 µg/mL) e carbenicilina (100 µg/mL). As bactérias foram incubadas a 37°C por 16 horas. As colônias isoladas foram, então, inoculadas em 10mL de caldo LB, contendo estreptomicina (50 µg/mL), com auxílio de ponteiras estéreis. As bactérias (pré-inóculo) contidas em erlenmeyers foram cultivadas a 37°C, durante 16 horas, sob agitação orbital constante, a 200 rpm (Certomat® IS, Sartorius). Após este tempo, 5mL do pré-inóculo foram transferidos para 500mL de meio YT (peptona de caseína 1,6%, extrato de levedura 1,0%, NaCl 0,5%, pH 7,2), 2X, onde as células bacterianas foram, novamente, cultivadas sob agitação nas mesmas condições. Passado este tempo, as células foram incubadas em banho de gelo durante 30 minutos, em seguida, transferidas para tubos de centrífuga de 50mL, previamente, resfriados e submetidos à centrifugação a 6.000 x g, a 4°C, por 10 minutos (Centrífuga Sorvall RC5B Plus, Rotor SS-34). O sobrenadante foi descartado e o precipitado celular lavado três vezes com glicerol 10% gelado, sendo que os volumes de glicerol utilizados foram de 40mL, 40mL e 30mL, respectivamente. As culturas foram centrifugadas nas mesmas condições anteriores, após cada lavagem. Depois de lavadas, as células

bacterianas foram ressuspensas, suavemente, em 200 $\mu$ L de meio líquido GYT (glicerol 10%, extrato de levedura 0,125%, triptona 0,25%, pH 7,0). Alíquotas de 50 $\mu$ L de células eletrocompetentes foram distribuídas em microtubos (gelados) estéreis de 1,5mL e armazenadas a -80°C, até a etapa de eletroporação. O controle positivo do método foi obtido através de uma reação de transformação de bactérias competentes, utilizando uma solução contendo plasmídeos circulares (pGEM-T) purificados e de concentração, previamente, estabelecida.

Alíquotas de 50 $\mu$ L das células TOP10 F' e 10 $\mu$ L de cada ligação foram usadas para a transformação das células bacterianas competentes. Os componentes foram homogeneizados, suavemente, e incubados no gelo por 10 minutos. Em seguida, as suspensões bacterianas foram transferidas para cubetas de 2mm e eletroporadas (Electroporator 2510, Eppendorf) sob tensão de 2.500V, durante 15 segundos. Imediatamente após, foi adicionado às cubetas 1mL de meio SOC [1mL de meio SOB (triptona 2%, m/v; extrato de levedura 0,5%, m/v; cloreto de sódio 10mM, cloreto de potássio 2,5mM, cloreto de magnésio 10mM, glucose 20mM, pH 7,0), 40 $\mu$ L de glicose 20%, 40 $\mu$ L de cloreto de magnésio 1M], previamente aquecido a 37°C, e homogeneizado, suavemente. As suspensões foram, então, transferidas para tubos falcon (15mL) estéreis e cultivadas sob agitação orbital a 150 rpm, durante 1 hora a 37°C.

Alíquotas de 50 $\mu$ L da cultura de bactérias foram plaqueadas em meio LB ágar, contendo estreptomicina (50 $\mu$ L/mL), carbenicilina (100 $\mu$ L/mL), IPTG (100 $\mu$ L/mL) e X-GAL (80 $\mu$ L/mL) e incubadas a 37°C, durante 16 horas. Somente as colônias transformadas, visualmente, de cor branca, foram cultivadas em 5mL de caldo LB, contendo estreptomicina (100 $\mu$ g/mL) e carbenicilina (50 $\mu$ g/mL), sob agitação a 180rpm, por 12 horas a 37°C.

## **6.10 Extração de DNA plasmidial**

Os plasmídeos transformados foram extraídos e purificados do cultivo bacteriano usando o *kit* NoLid para extração de DNA plasmidial da Macherey-Nagel, após a transformação bacteriana ser realizada. Para tanto, foram centrifugados 4mL de cada cultura selecionada, durante 30 segundos a 11.000 x g, à temperatura ambiente (Eppendorf® MiniSpin, Model 533). O sobrenadante foi descartado. O pellet foi ressuspensado em 250 $\mu$ L de tampão A1 e homogeneizado com o uso de um vortex. Em seguida, 250 $\mu$ L de tampão de lise A2 foram adicionados, invertidos cuidadosamente, e incubados por 5 minutos à temperatura ambiente. Uma quantidade de 350 $\mu$ L de tampão A3 foi adicionada e os

microtubos foram suavemente invertidos (6 e 8 vezes). As amostras foram centrifugadas a 11.000 x g, durante 5 minutos, à temperatura ambiente. Posteriormente, as amostras foram transferidas para colunas [NucleoSpin® Plasmid / Plasmid (NoLid)], previamente acopladas a microtubos de 2mL, e centrifugadas a 11.000 x g, durante 1 minuto, à temperatura ambiente. O sobrenadante foi descartado. Logo após esta etapa, 600µL de tampão de dessalinização A4 (que continha etanol 100%) foram adicionados às colunas, contendo as amostras de plasmídeos e estas foram centrifugadas nas mesmas condições anteriores. As colunas foram, então, recolocadas em seus devidos microtubos e centrifugadas a 11.000 x g por 2 minutos à temperatura ambiente. O DNA plasmidial foi eluído com tampão AE (em aproximadamente 50µL). Em seguida, as amostras foram armazenadas a -20°C para uso posterior.

### **6.11 Digestão enzimática com PvuII**

A comprovação da eficiência da transformação foi realizada através da digestão do DNA plasmidial utilizando-se a enzima PvuII. Para tanto, foram misturados 11,5µL de água livre de nuclease, 2µL de tampão G 10 X, 6µL de DNA (1µg) e 0,5µL de PvuII (10U) dentro de um microtubo estéril. O material a ser digerido foi submetido a uma temperatura de 37°C por 16 horas, após uma leve homogeneização e centrifugação. Posteriormente, o padrão de bandas foi analisado por eletroforese em gel de agarose 0,8%.

### **6.12 Sequenciamento nucleotídico**

Esta análise foi realizada, utilizando por base o método de terminação de cadeia (método de Sanger), através do sequenciador AppliedBiosystems 3730xl DNA Analyzer. O processo de sequenciamento foi feito, inteiramente, pela Macrogen Inc. Os iniciadores “universais” T7 promoter (5'-TAATACGACTCA CTATAGGG-3') e SP6 downstream (5'-ATTTAGGTGACACTATAG- 3') foram empregados para a amplificação dos reads de cada um dos clones.

### **6.13 Eletroforese em gel de poliacrilamida**

O perfil proteico da farinha das sementes de *S. grandiflora* foi realizado por eletroforese em gel de poliacrilamida (SDS – PAGE), onde os géis de concentração e separação foram obtidos a partir de uma solução estoque de acrilamida a 30% p/v e de N-N'-

metilenobis-acrilamida 0,8% p/v (LAEMMLI, 1970). O gel de concentração a 5% p/v foi preparado em tampão Tris-HCl 0,5M, pH 6,8, e o gel de separação 15% p/v, em tampão Tris-HCl 1,5M, pH 8,8, sendo acrescentado, em ambos, SDS 20% p/v. A polimerização foi conseguida pela adição de TEMED e PSA 10% p/v.

A amostra proteica (20 $\mu$ g) foi dissolvida em tampão Tris-HCl 0,0625M, pH 6,0, contendo 1% p/v de SDS, 10% v/v de glicerol e 1% v/v de  $\beta$ -mercaptoetanol e, posteriormente, imersas em água em ebulição durante 10 minutos. A eletroforese foi realizada em tampão de corrida Tris-HCl 0,025M, glicina 0,192M e SDS 0,1%, a 120 volts, 15mA, durante 2 horas.

Após a eletroforese, os géis foram corados com Coomassie Brilliant Blue R-250 com ácido acético 0,1% v/v, metanol e água deionizada, na proporção de 1:4:5, durante 2 horas e, então, descorados em solução de ácido acético glacial, metanol e água deionizada, na proporção de 1:4:5.

#### **6.14 Sequenciamento da região N - terminal**

A banda referente à SgL foi transferida do gel de poli-acrilamida 15% para uma membrana de PVDF, segundo TOWBIN; STAEHELIN; GORDON, 1979, com adaptações, e a obtenção dos aminoácidos que compõem a região NH<sub>2</sub> -terminal da lectina de *S. grandiflora* foi realizada pelo sequenciador automático de proteínas (Shimadzu PPSQ-23A), baseado no princípio da degradação de Edman. Os derivados feniltioidantoína dos aminoácidos (PTH-aminoácidos) foram detectados a 269nm, após separação em coluna de fase reversa C18 (4,6 x 2,5mm), conduzida sob condições isocráticas, de acordo com as instruções do fabricante. A sequência de aminoácidos obtida foi submetida ao alinhamento automático através do sistema NCBI-BLAST.

#### **6.15 Espectrometria de massas**

A banda referente à lectina de *S. grandiflora* foi recortada do gel de poli-acrilamida 15% e, então, descorada com acetonitrila 50% em bicarbonato de amônia 25mM, pH 8,8 e, posteriormente, submetidas à lavagens em água deionizada estéril. Logo após, o gel foi desidratado com 200 $\mu$ L de acetonitrila 100%. A acetonitrila foi removida e o gel foi completamente seco por liofilização. As bandas do gel foram reduzidas com 100 $\mu$ L de DDT 65mM em bicarbonato de amônia 100mM, por 30 minutos, a 56°C, em banho-maria. Uma

alquilação do gel com 100µL de Iodoacetamida 200 mM em bicarbonato de amônio 100mM, por 30 minutos, à temperatura ambiente, foi realizada no escuro. Depois de removida a solução anterior, os pedaços de gel foram lavados com 200µL de bicarbonato de amônio 100mM, por 10 minutos. A solução anterior foi removida e os pedaços de gel, desidratados com 200µL de acetonitrila 100mM. Removida a última solução, os pedaços de gel foram reidratados com 200µL de bicarbonato de amônio 100mM, por 10 minutos. Posteriormente, os pedaços do gel foram desidratados por mais duas vezes com 200µL de acetonitrila 100% e secos por liofilização. Adicionou-se, então, 20µL de solução de tripsina diluída (25ng/µL) e incubada, durante 45 minutos, para que fosse dado início à digestão. Decorrida a incubação anterior, adicionou-se 50µL de bicarbonato de amônia 40mM em acetonitrila 10% e, então, realizou-se a incubação do material por 24 horas a 37°C. Após dado período de incubação, o material foi submetido a ultrassom por 10 minutos e, em seguida, colocado no vortex por 20 segundos e, então, toda a solução foi removida para um outro microtubo, limpo. Ao gel restante, adicionou-se 30µL de ácido fórmico 5% em acetonitrila 50%, quando, então, o material foi submetido a vortex, por 20 segundos e, em seguida, deixado por 15 minutos em repouso em temperatura ambiente, depois 2 minutos no ultrassom e, novamente, 20 segundos no vortex. Posteriormente, removeu-se a solução com os peptídeos, juntando-a a solução do tubo limpo. O passo anterior foi repetido e centrifugado por 30 segundos a 14.000rpm. As amostras foram concentradas até cerca de 10µL e, em seguida, aplicadas ao espectrômetro de massas através do sistema de UPLC-ESI-Q-TOF, em um aparelho da Waters (Synapt, Manchester, UK). O resultado obtido da espectrometria foi analisado pelo programa Mascot, a fim de identificar os peptídeos de lectinas.

### **6.16 Análises de bioinformática**

Os reads obtidos do sequenciamento foram, inicialmente, processados através dos programas Phred/Phrap/Consed (EWING; GREEN, 1998; EWING *et al.*, 1998; GORDON; ABAJIAN; GREEN, 1998), onde a porção da sequência referente ao plasmídeo pGEM -T foi eliminada e os contigs de cada clone foram montados. Os contigs foram analisados com o auxílio do programa BioEdit (HALL, 1999) e BLAST (ALTSCHUL *et al.*, 1990), além de outras ferramentas de bioinformática disponíveis *online*, que se encontram nos sites do National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/tools>) e do Swiss Institute of Bioinformatics (<http://www.expasy.ch/tools>), entre outros. Os alinhamentos múltiplos das sequências de aminoácidos e de DNA foram realizados com o programa

ClustalW2 (BROOKSBANK *et al.*, 2003) e, posteriormente, corrigidos manualmente. Para a obtenção da sequência deduzida de aminoácidos, foram utilizadas as ferramentas de tradução disponíveis no banco de dados público Expert Protein Analysis System (<http://www.expasy.org/tools/dna/html>).

As sequências deduzidas de aminoácidos, no formato fasta, foram comparadas com as proteínas já depositadas no banco de dados genômicos do NCBI (GenBank), através do programa BLASTp, que alinha sequências de proteínas fornecidas em relação às sequências de proteínas já catalogadas pelo GenBank (<http://www.ncbi.nlm.nih.gov/blast/Blast/cgi>).

### 6.17 Modelagem computacional

Os modelos tridimensionais (3D) para as sequências deduzidas de aminoácidos foram calculados através de modelagem por homologia estrutural, usando os programas M4T (FERNANDEZ-FUENTES *et al.*, 2007a, 2007b; RYKUNOV *et al.*, 2009) e Phyre 2.0 (KELLEY *et al.*, 2015). Os modelos gerados para a SgL foram validados pelo *software* MolProbity e os parâmetros utilizados foram escolhidos por default (CHEN *et al.*, 2010).

### 6.18 Docking molecular

O docking molecular foi realizado utilizando-se o *software* Molegro Virtual Docking, usando-se o método Moldock. A estrutura dos glicanos foi obtida em banco de dados PDB. O Moldock é baseado em uma pesquisa algorítmica que é combinada com a evolução diferencial e um algoritmo de predição de cavidades. Um procedimento de re-docking foi adicionado para aumentar a precisão de encaixe nos sítios de ligação detectados. O MolDock Score (MDS) foi calculado usando funções de pontuações. Inicialmente, foram realizadas 10 corridas com o GRID de resolução de 0,20Å, no qual o número máximo de interações foi de 2.000, bem como o tamanho da população de 500 e 10 poses. A energia de interação proteína-ligante foi expressa em uma unidade arbitrária na forma de MolDock score (MDS), onde a valores mais negativos refletem interações mais fortes. O MDS foi calculado conforme a equação abaixo,  $MDS = E_{inter} + E_{intra}$ , onde  $E_{inter}$  é a energia de interação entre proteína e ligante (THOMSEN; CHRISTENSEN, 2006).



$$E_{inter} = \sum_{i \in T_{igand}} \sum_{j \in T_{igand}} \left[ E_{PLP}(r_{ij}) + 332.0 \frac{q_i q_j}{4 r_{ij}^2} \right]$$

## 7 RESULTADOS

No transcorrer dos próximos parágrafos se encontram os resultados obtidos dos experimentos já descritos.

### 7.1 Extração do RNA total das sementes de *Swartzia grandiflora*

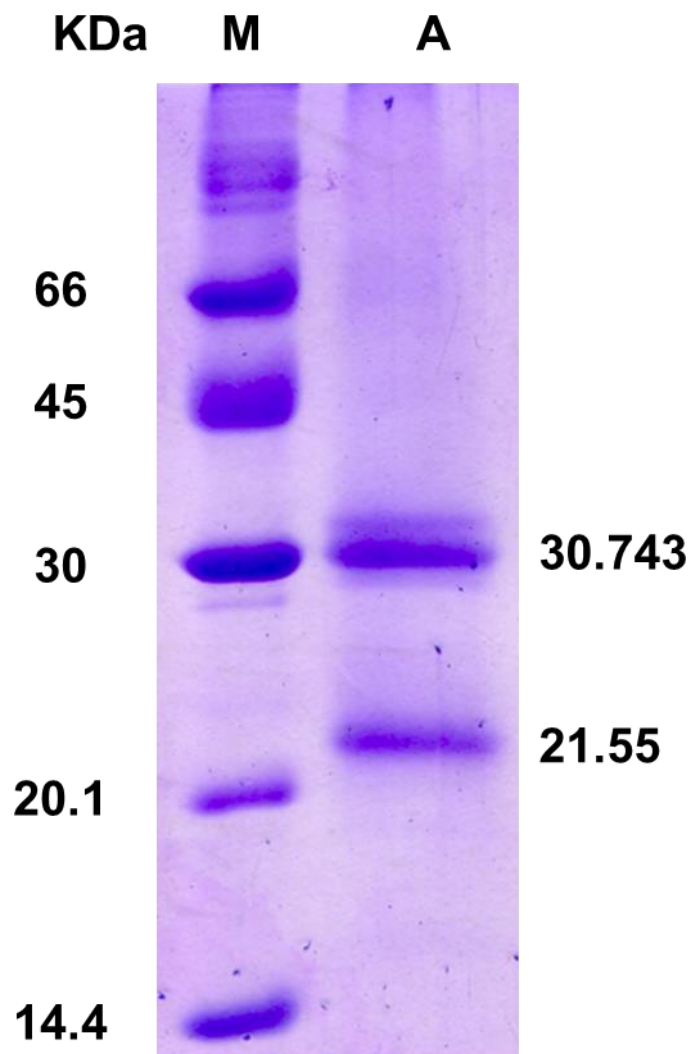
A concentração média de RNA das sementes de *Swartzia grandiflora* obtida foi de 1µg/µL. O perfil eletroforético para esta amostra de RNA demonstrou a presença de duas bandas (APÊNDICE A). A banda de maior tamanho possui 1,5kb, sendo referente ao RNA ribossomal 28S. Já a banda de tamanho igual a 1kb é referente ao RNA ribossomal 18S. Depois do tratamento com a enzima DNase, o RNA apresentou concentração média, aproximada, de 0,25µg/µL.

### 7.2 Amplificação rápida das extremidades do cDNA (RACE)

No sentido de se compreender as características do gene da SGL, esta pesquisa científica, inicialmente, planejou amplificar as regiões 5' UTR e 3' UTR, pela técnica RACE. Porém, a amplificação da região 5' UTR não foi eficiente, não havendo amplificação de nenhum produto de PCR. Na tentativa de obter a sequência inicial da lectina em estudo, a banda proteica de aproximadamente 30kDa (por ser o peso molecular característico das lectinas da família *Fabaceae*), vista após uma corrida eletroforética de 20µg da farinha das sementes de *Swartzia grandiflora* em gel de poliacrilamida 15% (Figura 6) foi transportada para uma membrana de PVDF e alguns resíduos de aminoácidos que compõem a região inicial da lectina foram obtidos por sequenciamento da região N-terminal, gerando um fragmento formado por 17 resíduos de aminoácidos, que a saber são: Thr, Lys, Phe, Val, Asp, Asn, Gln, Glu, Asp, Leu, Leu, Leu, Gln, Gly, Asp, Glu e Val. A similaridade desse fragmento com a de outras sequências do banco de proteínas não-redundantes, realizada por BLAST, foi demonstrada na Tabela 5. As maiores similaridades desse fragmento foram vistas, respectivamente, para as sequências de *Sophora japonica*, *Cladrastis kentukea*, *Robinia*

pseudoacacia e *Glycine max*. Adicionalmente, os dados da espectrometria de massas para os peptídeos, extraídos da banda de 30kDa, não foram satisfatórios, visto a ausência de peptídeos de lectinas quando feito a análise pelo programa Mascot.

Figura 6 – Eletroforese em gel de poliacrilamida 15% da farinha de sementes de *Swartzia grandiflora*. A letra M indica o poço contendo o marcador molecular Low Molecular Weight Calibration Kit for SDS Electrophoresis (GE Healthcare). A letra A indica o poço no qual foi aplicado 20 $\mu$ g da farinha de *Swartzia grandiflora* diluída em tampão da amostra contendo  $\beta$ -mercaptoetanol 1%



Fonte: Paulo Abraão (2015).

Tabela 5 – Sequências de aminoácidos com maior similaridade a região N-terminal da lectina de *Swartzia grandiflora* pelo programa BLASTp do NCBI, utilizando as sequências de proteínas não-redundantes como banco de dados (último acesso em 29 de agosto de 2015)

<b>Espécie</b>	<b>Cobertura (%)</b>	<b>E-value</b>	<b>Identidade (%)</b>	<b>N° de acesso</b>
<i>Sophora japonica</i>	83	5e-04	93	P93538.1
<i>Cladrastiskentukea</i>	88	0,002	81	Q39527.1
<i>Sophora japonica</i>	83	0,008	87	P93535.1
<i>Robinia pseudoacacia</i>	100	2,4	61	BAA36415.1
<i>Glycine max</i>	77	3,3	71	KHN10471.1

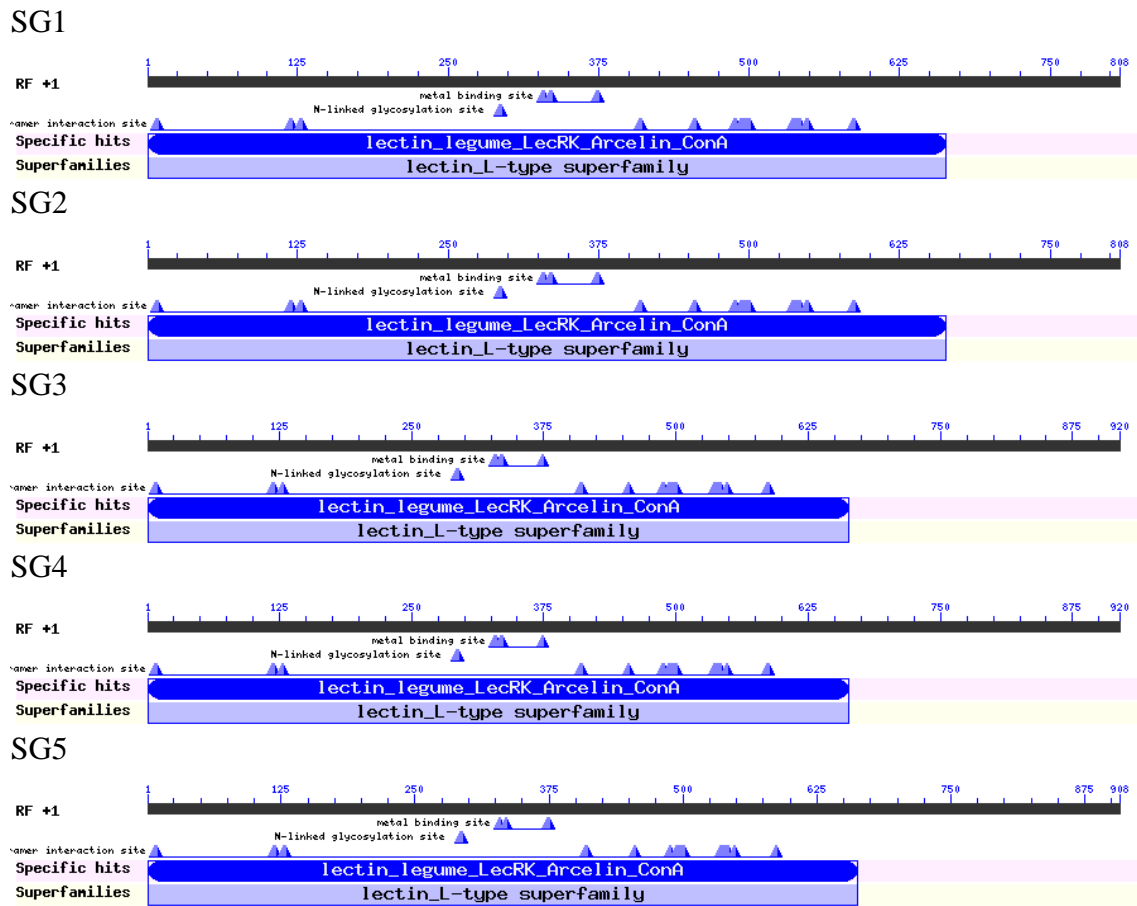
Fonte: elaborada pelo autor.

### 7.3 Caracterização in silico da região codificadora da SGL

Duas sequências nucleotídicas (reads) foram obtidas para cada um dos quinze clones recombinantes sequenciados, com a utilização dos iniciadores SP6 e T7promoter. Os contigs, sequências consenso dos reads, das 15 amostras foram montados com a ajuda do pacote Phred/Phrap/Consed, utilizado para leitura dos dados brutos em ambiente operacional Linux. Os resultados gerados mostraram quais dos contigs se apresentavam-se no sentido correto (5' → 3'). Para aqueles que se encontravam dispostos no sentido 3' → 5', foi necessário realizar a conversão da sequência para o sentido correto pelo programa Reverse Complement ([http://www.bioinformatics.org/sms/rev\\_comp.html](http://www.bioinformatics.org/sms/rev_comp.html)). Após este processo, foram obtidas quinze sequências nucleotídicas, denominadas da seguinte forma: SG1, SG2, SG3, SG4, SG5, SG6, SG6, SG7, SG8, SG9, SG11, SG12, SG13, SG14 e SG15. Posteriormente, todas estas sequências foram alinhadas como sequências contidas no banco de dados do NCBI pelo algoritmo do programa BLASTx, que utiliza uma sequência de nucleotídeos como fonte de busca para proteínas similares.

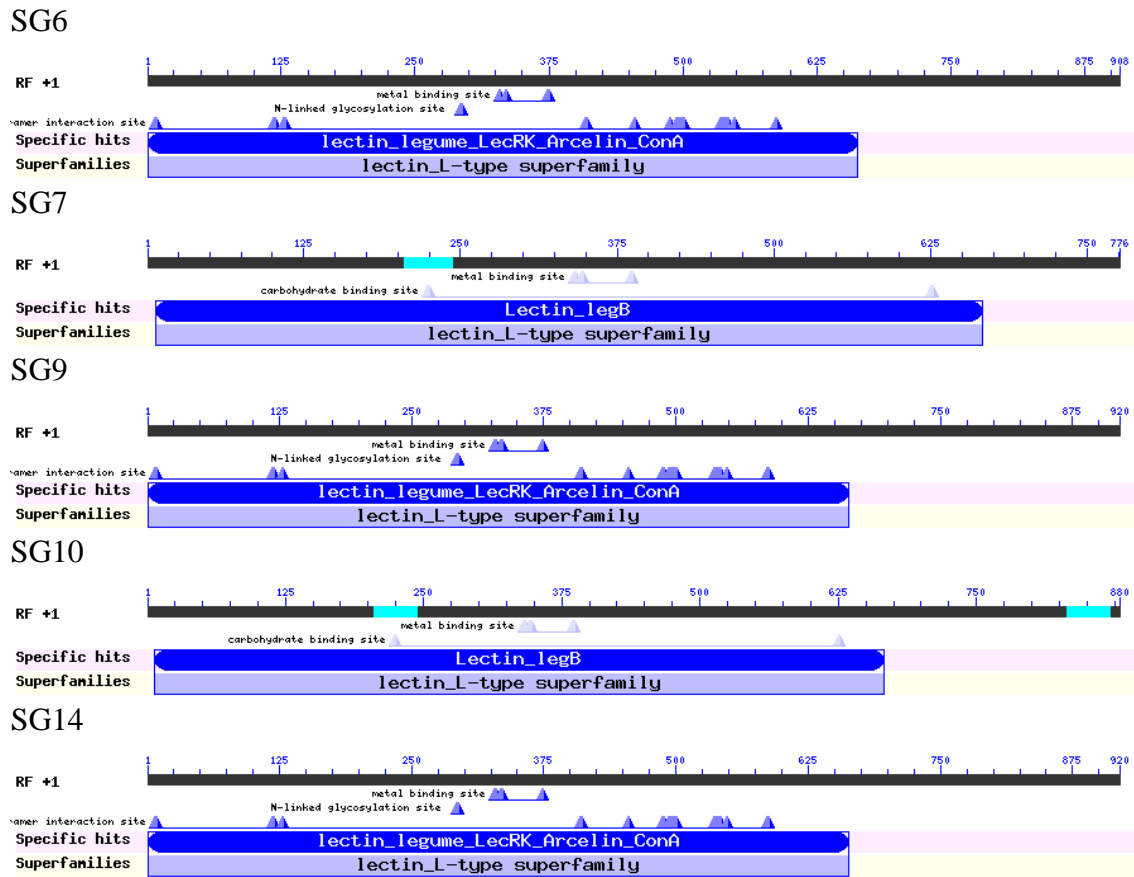
As sequências que não apresentaram similaridades com a classe das lectinas (SG8, SG11, SG12, SG13 e SG15) não foram utilizadas em análises posteriores deste estudo. Para as sequências restantes, foi visto uma grande similaridade com o domínio das lectinas de leguminosas, quando submetidas ao banco de domínios do NCBI (Figura 7 e Figura 8).

Figura 7 – Identificação dos domínios conservados das sequências pelo banco de dados do NCBI. O domínio encontrado para cada uma das sequências está descrito na tarja azul. O valor +1 indica o frame de leitura da sequência



Fonte: Fornecida pelo autor.

Figura 8 – Identificação dos domínios conservados das sequências pelo banco de dados do NCBI. O domínio encontrado para cada uma das sequências está descrito na tarja azul. O valor +1 indica o frame de leitura da sequência



Fonte: Fornecida pelo autor.

Uma matriz de diferença foi gerada com a intenção de descobrir se as sequências obtidas eram redundantes entre si (Tabela 6). A partir destes resultados observamos a presença de seis sequências nucleotídicas redundantes, denominadas da seguinte forma: SGL-A (referentes a SG1 e SG2), SGL-B (referentes a SG3 e SG4), SGL-C (referentes a SG5 e SG6), SGL-D (referentes a SG9 e SG14), SGL-E (referente a SG7) e SGL-F (referente a SG10). Para uma melhor interpretação deste resultado, na referida tabela, as sequências que compõem o grupo SGL-A foram tarjadas de amarelo; as do grupo SGL-B, de verde; as do grupo SGL-C, de azul; as do grupo SGL-D, de vermelho; a SGL-E, de roxo e a SGL-F, de rosa.

Tabela 6 – Matriz de diferença, realizada pelo programa BioEdit, das sequências nucleotídicas que codificam lectinas da semente de *S. grandiflora*

	SG1	SG2	SG3	SG4	SG5	SG6	SG7	SG9	SG10	SG14
SG1	ID	0	131	131	114	114	521	130	580	130
SG2	0	ID	131	131	114	114	521	130	580	130
SG3	131	131	ID	0	29	29	634	11	597	11
SG4	131	131	0	ID	29	29	634	11	597	11
SG5	114	114	29	29	ID	0	617	28	581	28
SG6	114	114	29	29	0	ID	617	28	581	28
SG7	521	521	634	634	617	617	ID	635	117	635
SG9	130	130	11	11	28	28	635	ID	598	0
SG10	580	580	597	597	581	581	117	598	ID	598
SG14	130	130	11	11	28	28	635	0	598	ID

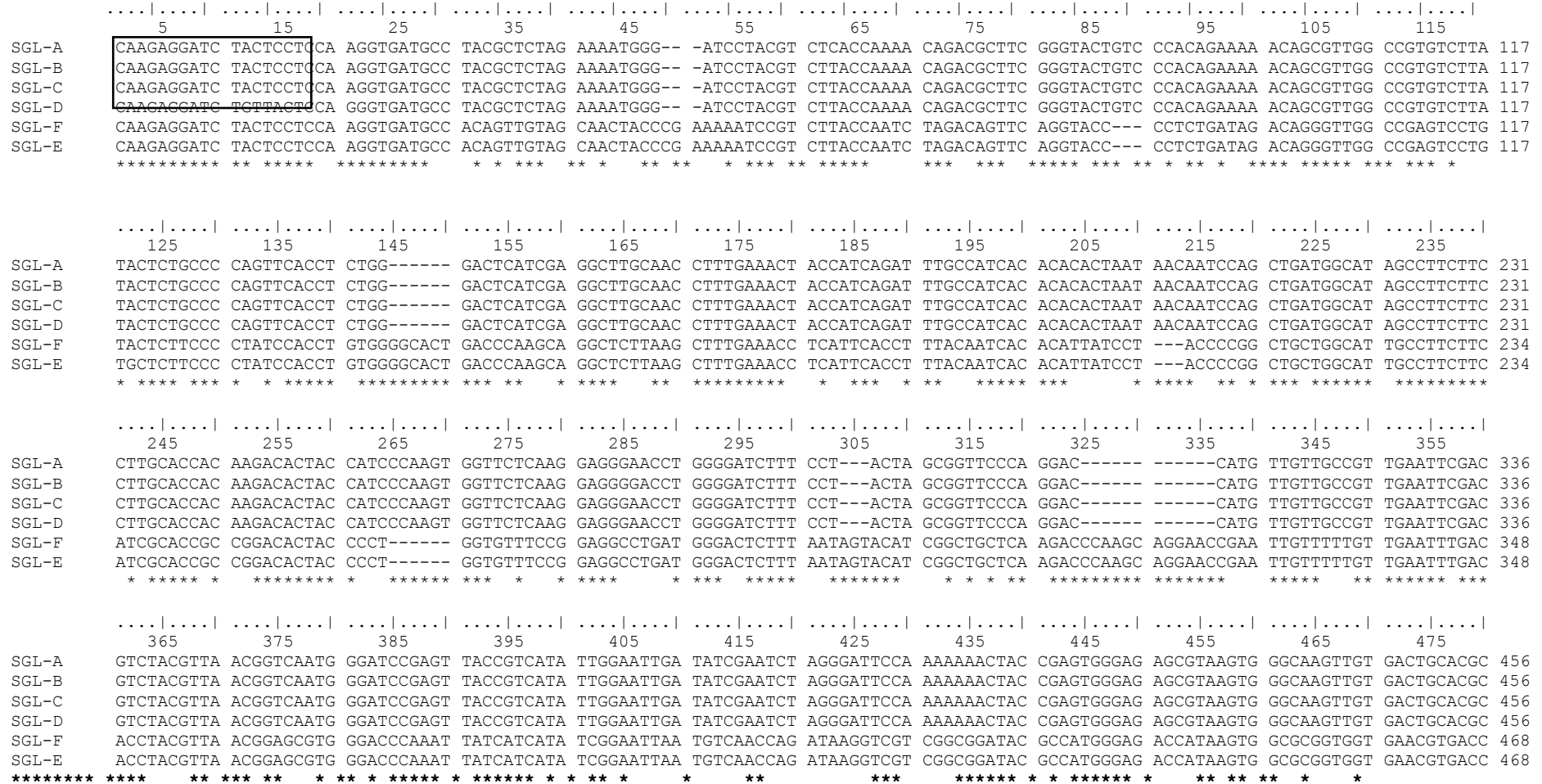
Fonte: elaborada pelo autor.

As sequências destes seis grupos foram alinhadas pelo ClustalW para melhor visualizar a região em que o iniciador ligou a sequência (demarcada com retângulo) e as regiões conservadas (demarcadas com \*), resultando na Figura 9. Em seguida, todas as sequências de nucleotídeos foram convertidas em aminoácidos pelo software livre Translate Tool (<http://web.expasy.org/translate/>), resultando em seis sequências de aminoácidos (Figura 10 - 15). A partir dessas análises, foi possível determinar, com exatidão, o códon de terminação e o sítio de poliadenilação para cada uma das seis sequências obtidas. Estas sequências, por conseguinte, foram alinhadas com o auxílio do software ClustalW, contido no programa BioEdit (Figura 16).





Figura 10 – Alinhamento múltiplo das sequências nucleotídicas da RACE 3’ que codificam para lectinas da *Swartzia grandiflora*. O número para o último nucleotídeo encontra-se descrito na margem direita do alinhamento. Os nucleotídeos conservados nas sequências alinhadas estão demarcados com um asterisco (\*)



Fonte: elaborada pelo autor.

Figura 11 – Sequência SGL-A, codificando uma lectina de *Swartzia grandiflora*. O iniciador utilizado na RACE 3' encontra-se sublinhado. O códon de parada está indicado por um asterisco (\*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente.

```

1  CAAGAGGATCTACTCCTCCAAGGTGATGCCTACGCTCTAGAAAATGGGATCCTACGTCTC
Q E D L LL Q G D A Y A L E N G I L R L 20

61 ACCAAAACAGACGCTTCGGGTACTGTCCCACAGAAAAACAGCGTTGGCCGTGTCTTATAC
T K T D A S G T V P Q K N S V G R V L Y 40

121 TCTGCCCCAGTTCACCTCTGGGACTCATCGAGGCTTGCAACCTTTGAACTACCATCAGA
S A P V H L W D S S R L A T F E T T I R 60

181 TTTGCCATCACACACACTAATAACAATCCAGCTGATGGCATAGCCTTCTTCCTTGCACCA
F A I T H T N NN P A D G I A F F L A P 80

241 CAAGACACTACCATCCCAAGTGGTTCTCAAGGAGGGAACCTGGGGATCTTTCCTACTAGC
Q D T T I P S G S Q G G N L G I F P T S 100

301 GGTTCCCAGGACCATGTTGTTGCCGTTGAATTCGACGTCTACGTTAACGGTCAATGGGAT
G S Q D H V V A V E F D V Y V N G Q W D 120

361 CCGAGTTACCGTCATATTGGAATTGATATCGAATCTAGGGATTCCAAAAAACTACCGAG
P S Y R H I G I D I E S R D S K K T T E 140

421 TGGGAGAGCGTAAGTGGGCAAGTTGTGACTGCACGCATAAACTATGACGCCAACCTTAAA
W E S V S G Q V V T A R I N Y D A N L K 160

481 AGGCTAAGTGTCTTCTGCTACTGCTTCTCCTCAAGGTCGTCGAGTCAATCAAGAACTTAC
R L S V S A T A S P Q G R R V N Q E T Y 180

541 ACTGTCACTTATGACAATGTTAACTTGAAGACTCTTCTTCTGAATATGTTAGAGTTGGA
T V T Y D N V N L K T L L P E Y V R V G 200

601 ATCTCTGGTGCCACAGGACAAGAGGTTGCCATACATGACATTCGCTCCTGGTATTTCTCT
I S G A T G Q E V A I H D I R S W Y F S 220

661 TCATACCTGGTGAATATCGCCAAGAAGGAGAAGGAGGACATGTATATGGCACAGTATGTG
S Y L V N I A K K E K E D M Y M A Q Y V 240

721 TGATTTTGGATCAAGGACTTCAATTTCCATCTATCTCTGTGATATACCAATAAGGACAAG
*
781 TGTGCTCTATAGTCTAAAAAAAAAAAAA

```

Fonte: elaborada pelo autor.

Figura 12 – Sequência SGL-B, codificando uma lectina de *Swartzia grandiflora*. O iniciador utilizado na RACE 3' encontra-se sublinhado. O códon de parada está indicado por um asterisco (\*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente

```

1  CAAGAGGATCTACTCCTCCAAGGTGATGCCTACGCTCTAGAAAATGGGATCCTACGTCTT
Q E D L LL Q G D A Y A L E N G I L R L 20

61 ACCAAAACAGACGCTTTCGGGTACTGTCCCACAGAAAAACAGCGTTGGCCGTGTCTTATAC
T K T D A S G T V P Q K N S V G R V L Y 40

121 TCTGCCCCAGTTCACCTCTGGGACTCATCGAGGCTTGCAACCTTTGAACTACCATCAGA
S A P V H L W D S S R L A T F E T T I R 60

181 TTTGCCATCACACACACTAATAACAATCCAGCTGATGGCATAGCCTTCTTCCTTGCACCA
F A I T H T N NN P A D G I A F F L A P 80

241 CAAGACACTACCATCCCAAGTGGTTCTCAAGGAGGGGACCTGGGGATCTTTCCTACTAGC
Q D T T I P S G S Q G G D L G I F P T S 100

301 GGTTCACAGGACCATGTTGTTGCCGTTGAATTCGACGTCTACGTTAACGGTCAATGGGAT
G S Q D H V V A V E F D V Y V N G Q W D 120

361 CCGAGTTACCGTCATATTGGAATTGATATCGAATCTAGGGATTCCAAAAAACTACCGAG
P S Y R H I G I D I E S R D S K K T T E 140

421 TGGGAGAGCGTAAGTGGCAAGTTGTGACTGCACGCATAAACTATGACGCCAACCTTAAA
W E S V S G Q V V T A R I N Y D A N L K 160

481 AGGCTAAGTGTCTTCTGCTACTGCTTCTCCTCAAGGTCGTCGAGTCAATCAAGAACTTAC
R L S V S A T A S P Q G R R V N Q E T Y 180

541 ACTGTCACCTTATGACAATGTTAACTTGAAGACTCTTCTTCTGAAATATGTTAGAGTTGGA
T V T Y D N V N L K T L L P E Y V R V G 200

601 ATCTCTGGTGCCACAGGACAAGAGGTTGCCATACATGACATTCGCTCCTGGTATTTCTTT
I S G A T G Q E V A I H D I R S W Y F F 220

661 TCATACCTGGTGAACATCGCCAAGAAGGAGAAGGAGGACATGTATATTGCACAGTATGTA
S Y L V N I A K K E K E D M Y I A Q Y V 240

721 TGATTTTGGATCAAGGACTTCAATTTCCATCTATCTCTGTGATATACCAATAAGGACAAG
*

781 TGTGCTCTATAGTCCATTGTGTTGTTGCTAACAAACGTCCCATCTTATGTTAATTATATGG
841 CCATCTATGGTCTATGTACAAAACCTTTTAAGCATGAGAGATAAGATGTATCGATTAAGT
901 GGTTGCGTAAAAAAAAAAAA

```

Fonte: elaborada pelo autor.

Figura 13 – Sequência SGL-C, codificando uma lectina de *Swartzia grandiflora*. O iniciador utilizado na RACE 3' encontra-se sublinhado. O códon de parada está indicado por um asterisco (\*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente

```

1  CAAGAGGATCTACTCCTCCAAGGTGATGCCTACGCTCTAGAAAATGGGATCCTACGTCTT
Q E D L LL Q G D A Y A L E N G I L R L 20
61 ACCAAAACAGACGCTTCGGGTACTGTCCCACAGAAAAACAGCGTTGGCCGTGTCTTATAC
T K T D A S G T V P Q K N S V G R V L Y 40
121 TCTGCCCCAGTTACCTCTGGGACTCATCGAGGCTTGCAACCTTTGAACTACCATCAGA
S A P V H L W D S S R L A T F E T T I R 60
181 TTTGCCATCACACACACTAATAACAATCCAGCTGATGGCATAGCCTTCTTCTTGCACCA
F A I T H T N NN P A D G I A F F L A P 80
241 CAAGACACTACCATCCCAAGTGGTTCTCAAGGAGGGAACCTGGGGATCTTTCCTACTAGC
Q D T T I P S G S Q G G N L G I F P T S 100
301 GGTTCACAGGACCATGTTGTTGCCGTTGAATTCGACGTCTACGTTAACGGTCAATGGGAT
G S Q D H V V A V E F D V Y V N G Q W D 120
361 CCGAGTTACCGTCATATTGGAATTGATATCGAATCTAGGGATTCCAAAAAACTACCGAG
P S Y R H I G I D I E S R D S K K T T E 140
421 TGGGAGAGCGTAAGTGGGCAAGTTGTGACTGCACGCATAAACTATGACGCCAACTTTAAA
W E S V S G Q V V T A R I N Y D A N F K 160
481 AGGCTAAGTGTTTTCTGCTACTGCTTCTCCTCAAGGTCGTCGAGTCAATCAAGAACTTAC
R L S V S A T A S P Q G R R V N Q E T Y 180
541 ACTGTCACCTTATGACAATGTAACTTGAAGACTCTTCTTCTGAATATGTTAGAGTTGGA
T V T Y D N V N L K T L L P E Y V R V G 200
601 ATCTCTGGTGCCACAGGACAAGAGGTTGCCATACATGACATTCGCTCCTGGTATTTCTCT
I S G A T G Q E V A I H D I R S W Y F S 220
661 TCATACCTGGTGAACATCGCCAAGAAGGAGAAGGAGGACATGTATATTGCACAGTATGTG
S Y L V N I A K K E K E D M Y I A Q Y V 240
721 TGATTTTGGATCAAGGACTTCAATTTCCATCTATCTCTGTGATATAACCAATAAGGACAAG
*
781 TGTGCTCTATAGTCCATTGTGTTGTTGCTAACACGTCCTCATCTTATGTTAATCATATGG
841 CCATCTATGGTCTATGTACCAAACTTTTAAGCATGAGAGATAAGATGTATCGAAAAAAA
901 AAAAA

```

Fonte: elaborada pelo autor.

Figura 14 – Sequência *sgl-d*, codificando uma lectina de *Swartzia grandiflora*. O iniciador utilizado na *race 3'* encontra-se sublinhado. O códon de parada está indicado por um asterisco (\*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente

```

1   CAAGAGGATCTGTTACTCCAGGGTGATGCCTACGCTCTAGAAAATGGGATCCTACGTCTT
Q E D L LL Q G D A Y A L E N G I L R L 20

61  ACCAAAACAGACGCTTCGGGTACTGTCCCACAGAAAAACAGCGTTGGCCGTGTCTTATAC
T K T D A S G T V P Q K N S V G R V L Y 40

121 TCTGCCCCAGTTCACCTCTGGGACTCATCGAGGCTTGCAACCTTTGAAACTACCATCAGA
S A P V H L W D S S R L A T F E T T I R 60

181 TTTGCCATCACACACACTAATAACAATCCAGCTGATGGCATAGCCTTCTTCCTTGCACCA
F A I T H T N NN P A D G I A F F L A P 80

241 CAAGACACTACCATCCCAAGTGGTTCTCAAGGAGGGAACCTGGGGATCTTTCCTACTAGC
Q D T T I P S G S Q G G N L G I F P T S 100

301 GGTTCCCAGGACCATGTTGTTGCCGTTGAATTCGACGTCTACGTTAACGGTCAATGGGAT
G S Q D H V V A V E F D V Y V N G Q W D 120

361 CCGAGTTACCGTCATATTGGAATTGATATCGAATCTAGGGATTCCAAAAAACTACCGAG
P S Y R H I G I D I E S R D S K K T T E 140

421 TGGGAGAGCGTAAGTGGGCAAGTTGTGACTGCACGCATAAACTATGACGCCAACCTTAAA
W E S V S G Q V V T A R I N Y D A N L K 160

481 AGGCTAAGTGTTTCTGCTACTGCTTCTCCTCAAGGTCGTCGAGTCAATCAAGAACTTAC
R L S V S A T A S P Q G R R V N Q E T Y 180

541 ACTGTCACCTTATGACAATGTAACTTGAAGACTCTTCTTCTGAATATGTTAGAGTTGGA
T V T Y D N V N L K T L L P E Y V R V G 200

601 ATCTCTGGTGCCACAGGACAAGAGGTTGCCATACATGACATTCGCTCCTGGTATTTCTCT
I S G A T G Q E V A I H D I R S W Y F S 220

661 TCATACCTGGTGAACATCGCCAAGAAGGAGAAGGAGGACATGTATATTGCACAGTATGTG
S Y L V N I A K K E K E D M Y I A Q Y V 240

721 TGATTTTGGATCAAGGACTTCAATTTCCATCTATCTCTGTGATATACCAATAAGGACAAG
*

781 TGTGCTCTATAGTCCATTGTGTTGTTGCTAACAACGTCCCATCTTATGTTAATTATATGG
841 CCATCTATGGTCTATGTACCAAACTTTTAAGCATGAGAGATAAGATGTATCGATTAAGT
901 GGTTGCTAAAAA

```

Fonte: elaborada pelo autor.

Figura 15 – Sequência SGL-E, codificando uma lectina de *Swartzia grandiflora*. O iniciador utilizado na RACE 3' encontra-se sublinhado. O códon de parada está indicado por um asterisco (\*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente

```

1  CAAGAGGATCTACTCCTCCAAGGTGATGCCACAGTTGTAGCAACTACCCGAAAAATCCGT
Q E D L LL Q G D A T V V A T T R K I R 20

61  CTTACCAATCTAGACAGTTCAGGTACCCCTCTGATAGACAGGGTTGGCCGAGTCCCTGTGC
L T N L D S S G T P L I D R V G R V L C 40

121 TCTTCCCCTATCCACCTGTGGGGCACTGACCCAAGCAGGCTCTTAAGCTTTGAAACCTCA
S S P I H L W G T D P S R L L S F E T S 60

181 TTCACCTTTACAATCACACATTATCCTACCCCGGCTGCTGGCATTGCCTTCTTCATCGCA
F T F T I T H Y P T P A A G I A F F I A 80

241 CCGCCGGACACTACCCCTGGTGTTCGGGAGGCCTGATGGGACTCTTTAATAGTACATCG
P P D T T P G V S G G L M G L F N S T S 100

301 GCTGCTCAAGACCCAAGCAGGAACCGAATTGTTTTGTTGAATTTGACACCTACGTTAAC
A A Q D P S R N R I V F V E F D T Y V N 120

361 GGAGCGTGGGACCCAAATTATCATCATATCGGAATTAATGTCAACCAGATAAGGTCGTCG
G A W D P N Y H H I G I N V N Q I R S S 140

421 GCGGATACGCCATGGGAGACCATAAGTGGCGGGTGGTGAACGTGACCATAAGCTATGAC
A D T P W E T I S G A V V N V T I S Y D 160

481 AGTGAATCTAAGAAGCTAAGTGTTCCTCTTTTATCCGTGTTATCCATGCACTCATCAA
S E S K K L S V S S F Y P C Y P C T H Q 180

541 GGACCTTACACCGTGAGTCTCGACGTTGACTTGTCGGAAGTCCGTCTTCCGGAATGGGTT
G P Y T V S L D V D L S E V R L P E W V 200

601 AGAATTGGAATCTCGGGTACCAACGGAGCAAACAGGGCAAACCTATGAGATTCTCTCATGG
R I G I S G T N G A N R A N Y E I L S W 220

661 TCTTTCAAGTCAACCTTAAAGAGCAACGCTGTCAACGAGGACAAGGAGAACATGCGTATT
S F K S T L K S N A V N E D K E N M R I 240

721 GCGCAATATGTGTGATCATCATATATATGTACTACCATATAATCCCGCGGCCA
A Q Y V *

```

Fonte: elaborada pelo autor.

Figura 16 – Sequência SGL-F, codificando uma lectina de *Swartzia grandiflora*. O iniciador utilizado na RACE 3' encontra-se sublinhado. O códon de parada está indicado por um asterisco (\*). O sítio de poliadenilação do peptídeo está sublinhado por uma linha sólida. Os números para o primeiro nucleotídeo e o último resíduo de aminoácido estão marcados em cada linha, na esquerda e na direita, respectivamente

```

1  CAAGAGGATCTACTCCTCCAAGGTGATGCCACAGTTGTAGCAACTACCCGAAAAATCCGT
Q E D L LL Q G D A T V V A T T R K I R 20

61  CTTACCAATCTAGACAGTTCAGGTACCCCTCTGATAGACAGGGTTGGCCGAGTCCGTGAC
L T N L D S S G T P L I D R V G R V L Y 40

121 TCTTCCCCTATCCACCTGTGGGGCACTGACCCAAGCAGGCTCTTAAGCTTTGAAACCTCA
S S P I H L W G T D P S R L L S F E T S 60

181 TTCACCTTTACAATCACACATTATCCTACCCCGGCTGCTGGCATTGCCTTCTTCATCGCA
F T F T I T H Y P T P A A G I A F F I A 80

241 CCGCCGGACACTACCCCTGGTGTTCGGGAGGCCTGATGGGACTCTTTAATAGTACATCG
P P D T T P G V S G G L M G L F N S T S 100

301 GCTGCTCAAGACCCAAGCAGGAACCGAATTGTTTTGTTGAATTTGACACCTACGTTAAC
A A Q D P S R N R I V F V E F D T Y V N 120

361 GGAGCGTGGGACCCAAATTATCATCATATCGGAATTAATGTCAACCAGATAAGGTCGTCG
G A W D P N Y H H I G I N V N Q I R S S 140

421 GCGGATACGCCATGGGAGACCATAAGTGGCGGGTGGTGAACGTGACCATAAGCTATGAC
A D T P W E T I S G A V V N V T I S Y D 160

481 AGTGAATCTAAGAAGCTAAGTGTTCCTCTTTTATCCGTGTTATCCATGCACTCATCAA
S E S K K L S V S S F Y P C Y P C T H Q 180

541 GGACCTTACACCGTGAGTCTCGACGTTGACTTGTCGGAAGTCCGTCTTCCGGAATGGGTT
G P Y T V S L D V D L S E V R L P E W V 200

601 AGAATTGGAATCTCGGGTACCAACGGAGCAAACAGGGCAAACCTATGAGATTCTCTCATGG
R I G I S G T N G A N R A N Y E I L S W 220

661 TCTTTCAAGTCAACCTTAAAGAGCAACGCTGTCAACGAGGACAAGGAGAACATGCGTATT
S F K S T L K S N A V N E D K E N M R I 240

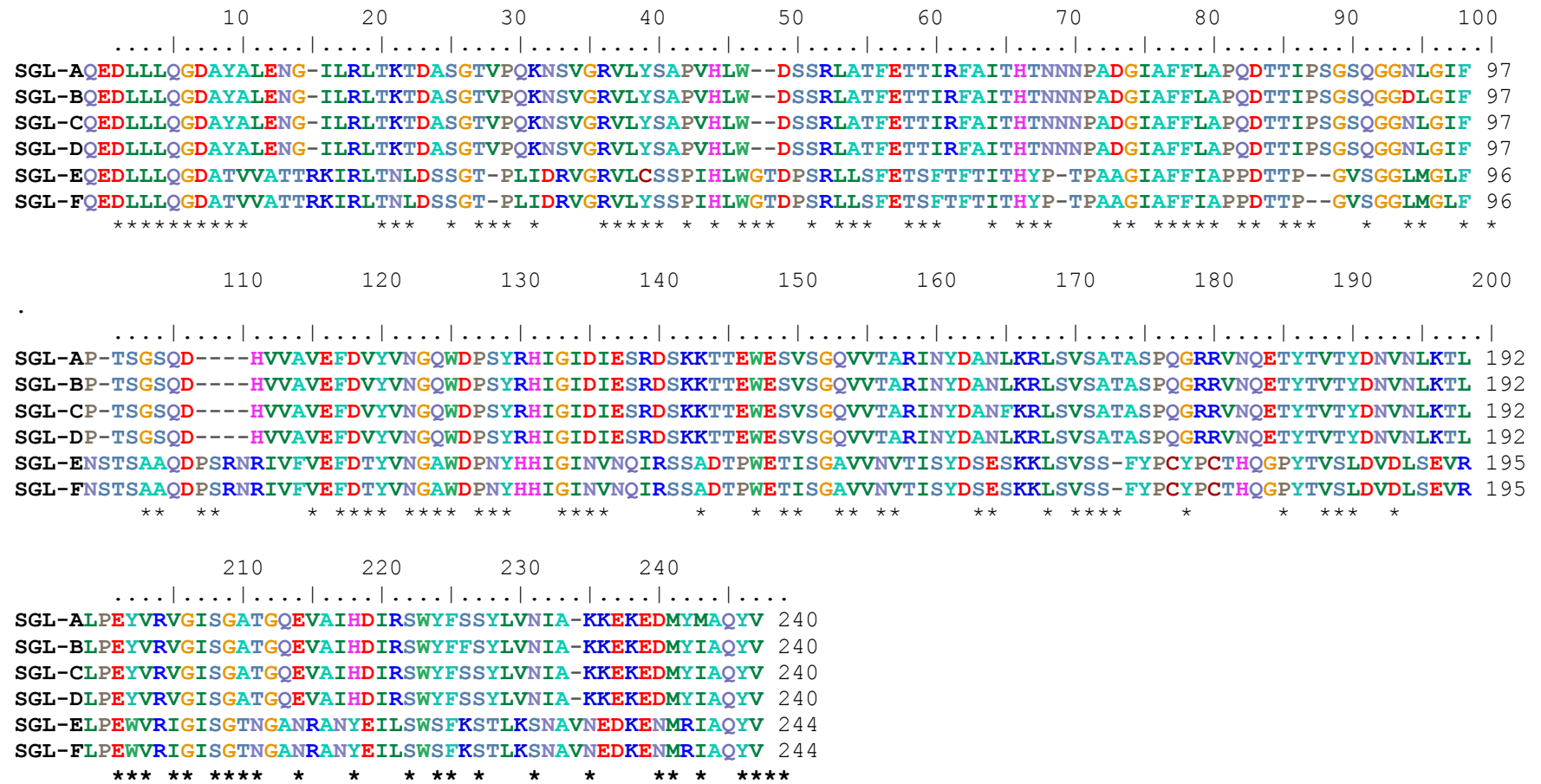
721 GCGCAATATGTGTGATCATCATATATATGTACTACCATATGCCATATACATCCCAATAAG
A Q Y V *

781 GACTATGTTCCACTAAGTGTGTTGCATATGTACTATACATATGGATATCATATTATAAAC
841 TAATTATTATAAAGATTAAGCTAGCAAAAAAAAAA

```

Fonte: elaborada pelo autor.

Figura 17 – Alinhamento múltiplo das sequências de aminoácidos que codificam para lectinas da *Swartzia grandiflora*. O número para o último resíduo de aminoácido encontra-se descrito na margem direita do alinhamento. Os resíduos de aminoácidos conservados nas sequências alinhadas estão demarcados com um asterisco (\*)



Fonte: elaborada pelo autor.



As análises comparativas, do tamanho das sequências obtidas nesta pesquisa, demonstram que as sequências SGL-A, SGL-B, SGL-C e SGL-D possuem 240 aminoácidos. Enquanto as sequências da SGL-E e SGL-F são compostas por 244 aminoácidos. O peso molecular e *pI*, para as seis sequências em estudo, estimado pelo programa contido em [web.expasy.org/cgi-bin/compute\\_pi/pi\\_tool](http://web.expasy.org/cgi-bin/compute_pi/pi_tool) foi, em média, 26.675,55 Da, de peso molecular e 5,328 de *pI*. Os valores individuais destas sequências estão descritos na Tabela 7.

Tabela 7 – Peso molecular e ponto isoelétrico estimado para as sequências de aminoácidos deduzidas obtidos da RACE 3' da *S. grandiflora*

<b>Sequência</b>	<b>Peso molecular (Da)</b>	<b><i>pI</i></b>
SGL-A	26587.58	5.37
SGL-B	26630.63	5.22
SGL-C	26603.56	5,37
SGL-D	26569.54	5.37
SGL-E	26800.99	5.32
SGL-F	26861.03	5.32

Fonte: elaborada pelo autor.

Uma comparação dessas seis sequências de aminoácidos de lectinas de *S. grandiflora* (SGL), codificados pelos fragmentos de cDNA clonados, foi realizada através de uma matriz de identidade e uma matriz de diferença pelo *BioEdit*, podendo ser visualizada na Tabela 8. Esse resultado indica a presença de dois grupos de sequências similares. Um dos grupos é composto pelas sequências SGL-A, SGL-B, SGL-C e SGL-D e diferem entre si em, no máximo, três resíduos de aminoácidos, a saber: o Asp-93 da SGL-B (nas outras sequências é visto Asn), a Phe-167 da SGL-C (nas outras sequências é visto Leu) e Phe-220 (nas outras sequências é visto Ser). O outro grupo visto na Tabela 8, é formado pelas sequências SGL-E e SGL-F e diferem, entre si, apenas no resíduo de posição quarenta, onde na SGL-E é visto uma Cys, enquanto que na SGL-F é visto um Tyr. Sabendo dessas diferenças, a sequência deduzida de aminoácidos da SGL-A foi escolhida como representante para posteriores análises do grupo um, enquanto que para o grupo dois foi escolhida a SGL-E como representante.

Tabela 8 – Matriz de identidade (acima da diagonal) e matriz de diferença (abaixo da diagonal) para cada par de sequências comparadas. Estes números foram calculados com base no alinhamento múltiplo de sequências, mostrado na figura 15

	SGL-A	SGL-B	SGL-C	SGL-D	SGL-E	SGL-F
SGL-A	-	0.987	0.991	0.995	0.527	0.531
SGL-B	3	-	0.987	0.991	0.531	0.536
SGL-C	2	3	-	0.995	0.531	0.536
SGL-D	1	2	1	-	0.531	0.536
SGL-E	111	110	110	110	-	0.995
SGL-F	110	109	109	109	1	-

Fonte: elaborada pelo autor.

As sequências deduzidas para a SGL-A e SGL-E, quando alinhadas pelo programa BLAST com o banco de proteínas não-redundantes e PDB (Protein Data Bank) do NCBI, apresentaram similaridade com as sequências que codificam lectinas e com as sequências de lectinas de estruturas tridimensionais já resolvidas. Como podemos ver na Tabela 9, a SGL-A apresentou alta similaridade com a lectina isolada de *Cladratiskentukea*, com lectinas do gênero *Sophorae* e com a lectina de *Apios americana*. A sequência da SGL-E apresentou similaridade com as lectinas de *Phaseolus vulgaris* *Sesbaniabispsonia* (Tabela 10), além das mesmas lectinas vistas para a SGL-A, com exceção da *Apios americana*. Estes resultados foram gerados pelo alinhamento de ambas as sequências com o banco de proteínas não-redundantes.

As similaridades mais expressivas do alinhamento da SGL-A e SGL-E, utilizando o PDB, foram dispostas nas Tabelas 11 e 12. As lectinas das espécies de *Platypodium elegans*, *Leucomphalos mildbraedii*, *Pterocarpos angolensis* e *Maackia amurensis* foram as que apresentaram maior identidade com a sequência da SGL-A. Já para a sequência da SGL-E foram *Leucomphalos mildbraedii*, *Platypodium elegans*, *Vatairea macrocarpa*, *Phaseolus vulgaris* e *Lablab purpureus*, respectivamente.

Tabela 9 – Sequências de aminoácidos com maior similaridade à estrutura primária da SGL-A (bit score > 200), pelo programa BLASTp, do NCBI, utilizando as sequências de proteínas não-redundantes como banco de dados (último acesso em 20 de agosto de 2015)

<b>Espécie</b>	<b>Cobertura (%)</b>	<b>Bit score</b>	<b>E-value</b>	<b>Identidade (%)</b>	<b>Nº de acesso</b>
<i>Cladrastiskentukea</i>	100	231	2e-71	51	Q39529
<i>Sophoraalopecuroides</i>	100	228	5e-70	49	AAY68291
<i>S. flavescens</i>	99	228	5e-70	50	ACD13798
<i>S. flavescens</i>	99	224	2e-68	50	AAG00508
<i>S. alopecuroides</i>	100	223	4e-68	49	ACD39390
<i>Apiosamericana</i>	98	209	9e-63	50	BAN37442

Fonte: elaborada pelo autor.

Tabela 10 – Sequências de aminoácidos com maior similaridade à estrutura primária da SGL-E (bit score > 186), pelo programa BLASTp, do NCBI, utilizando as sequências de proteínas não-redundantes como banco de dados (último acesso em 20 de agosto de 2015)

<b>Espécie</b>	<b>Cobertura (%)</b>	<b>Bit score</b>	<b>E-value</b>	<b>Identidade (%)</b>	<b>N° de acesso</b>
<i>Cladrastiskentukea</i>	100	225	8e-69	50	Q39529
<i>Sophoraalopecuroides</i>	95	197	6e-58	48	AAY68291
<i>S. alopecuroides</i>	100	191	7e-56	46	ACD39390
<i>Phaseolus vulgaris</i>	100	190	3e-55	45	XP_007143413
<i>Sesbaniabispinosa</i>	92	186	2e-54	46	ABX59683.1
<i>S. flavescens</i>	99	187	4e-54	47	AAG00508

Fonte: elaborada pelo autor.

Tabela 11 – Lectinas com estrutura tridimensional conhecida e com maior similaridade à estrutura primária da SGL-A (Bit score > 166), obtido pelo BLASTp, utilizando o banco Protein Data Bank (PDB) (último acesso em 20 de agosto de 2015)

<b>Espécie</b>	<b>Cobertura (%)</b>	<b>Bit score</b>	<b>E-value</b>	<b>Identidade (%)</b>	<b>N° de acesso</b>
<i>Platypodiumelegans</i>	99	196	7e-61	44	3ZVX_A
<i>Leucomphalosmildbraedii</i>	93	183	2e-56	47	2FMD_A
<i>Pterocarpusangolensis</i>	98	175	5e-53	43	1Q8O_A
<i>Pterocarpusangolensis</i>	98	175	6e-53	43	1N3O_A
<i>Maackiaamurensis</i>	92	167	6e-50	45	1DBN_A

Fonte: elaborada pelo autor.

Tabela 12 – Lectinas com estrutura tridimensional conhecida e com maior similaridade à estrutura primária da SGL-E (Bit score > 160), obtido pelo BLASTp, utilizando o banco Protein Data Bank (PDB) (último acesso em 20 de agosto de 2015)

<b>Espécie</b>	<b>Cobertura (%)</b>	<b>Bit score</b>	<b>E-value</b>	<b>Identidade (%)</b>	<b>Nº de acesso</b>
<i>Leucomphalosmildbraedii</i>	95	179	1e-54	47	2FMD_A
<i>Platypodiumelegans</i>	96	174	1e-52	42	3ZVX_A
<i>Vataireamacrocarpa</i>	93	171	2e-51	44	4U2A_A
<i>Phaseolus vulgaris</i>	95	162	5e-48	41	1FAT_A
<i>Lablab purpureus</i>	97	162	1e-47	40	3UJO_A

Fonte: elaborada pelo autor.

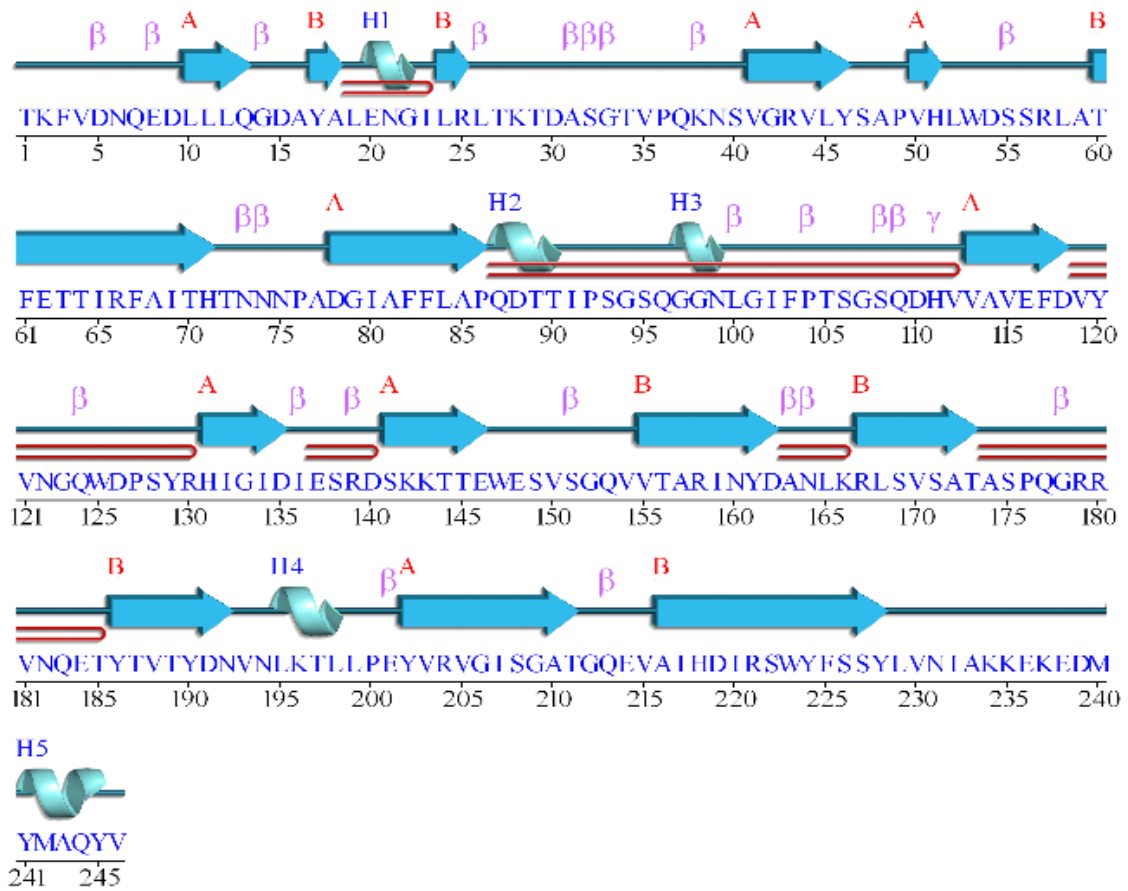
As estruturas secundárias da SGL-A e SGL-E, analisadas pelo *software* ProFunc (<https://www.ebi.ac.uk/thornton-srv/databases/profunc/>), indicam que as sequências estão organizadas, predominantemente, como folhas- $\beta$ . Esta análise apresentou, respectivamente, quinze e dezessete regiões preditas de folha- $\beta$  para a SGL-A e SGL-E (indicadas pelas setas amarelas das Figuras 17 e 18. Esta análise também evidencia a presença de poucas  $\alpha$ -hélices, cinco na SGL-A e apenas uma na SGL-E, como pode ser visto na topologia das estruturas, como mostra a Figura 19.

Os melhores modelos tridimensionais das sequências deduzidas da SGL-A e SGL-E foram deduzidos pelo programa M4T (Figura 20). Este programa utilizou como referência modelos proteicos de estruturas resolvidas que apresentaram maior similaridade com as sequências em estudo, por meio do algoritmo do próprio programa citado. As estruturas resolvidas das lectinas de *Lablab purpureus* (3UJQ\_A) e a de *Vatairea macrocarpa* foram utilizadas para a elaboração dos modelos das sequências em estudo. A validação dos modelos obtidos para a SGL-A e SGL-E foi realizada pelo *software* MolProbity, utilizando o gráfico de Ramachandran como parâmetro (Figuras 21 e 22). A estrutura obtida para a SGL-A não apresentou nenhum aminoácido fora da região delimitada permitida. Já a estrutura tridimensional gerada para a SGL-E apresentou três resíduos (His-67, Cys-174 e Tyr-175) fora das delimitações, mas esta pode ser considerada aceitável, pois o resultado apresentou mais de 95% de confiabilidade.

Os sítios de ligação a  $\text{Ca}^{2+}$  e  $\text{Mn}^{2+}$ , N-glicosilação, de formação de estruturas homodiméricas e homotetraméricas, tanto da SGL-A quanto da SGL-E, deduzidos pelo programa ConservadProtein Domain Family (presente no NCBI) foram representados, nas Figuras 23, 24 e 25, respectivamente.

Os resíduos de aminoácidos da SGL-A que interagem com  $\text{Ca}^{2+}$  e  $\text{Mn}^{2+}$  são Glu-110, Asp-112, Asn-116 e His-125. Já para a SGL-E, os resíduos responsáveis pela interação com estes cátions são Glu-114, Asp-116, Asn-120 e His-129. As medições das distâncias, realizadas pelo programa PyMol, entre os átomos que formam esses resíduos com os cátions e as moléculas de água foi de, aproximadamente, 2,2Å. O sítio de N-glicosilação predito das sequências de SGL-A e SGL-E é composto, por Pro-98 e Asn-97, respectivamente.

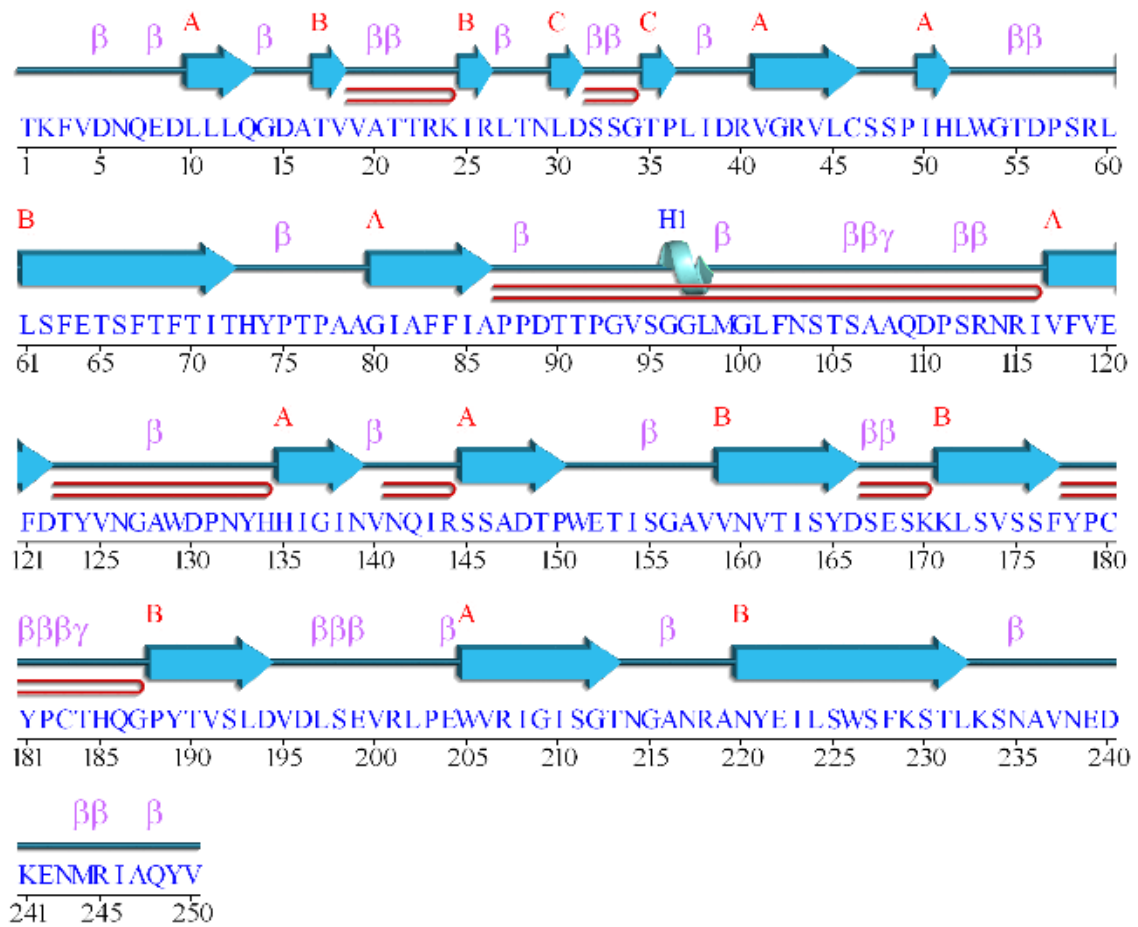
Figura 18 – Predição da estrutura secundária da SGL-A, com a região N-terminal inserida, pelo programa ProFunc



Fonte: elaborada pelo autor.



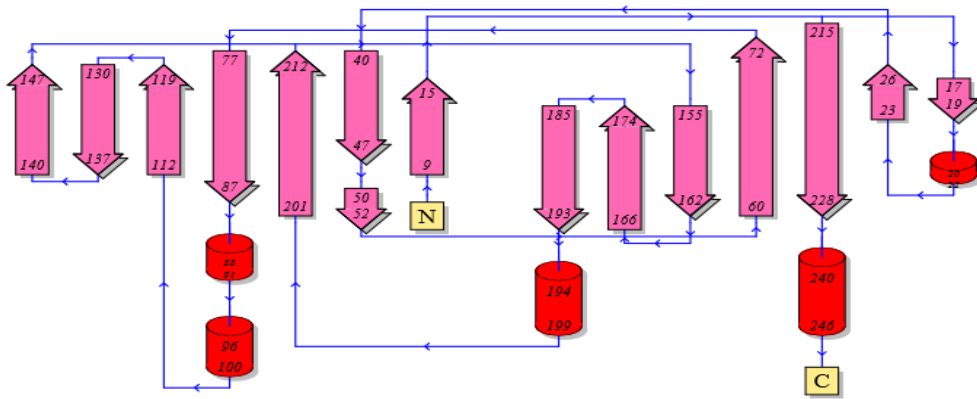
Figura 19 – Predição da estrutura secundária da SGL-E, com a região N-terminal inserida, pelo programa ProFunc



Fonte: elaborada pelo autor.

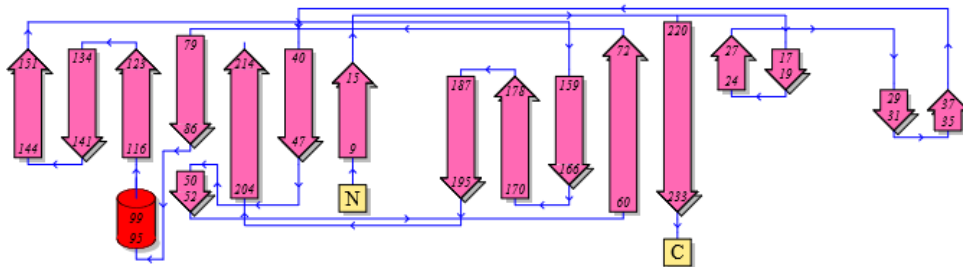
Figura 20 – Predição topológica da estrutura secundária da SGL-A e SGL-E, com a região N-terminal inserida, pelo programa ProFunc

### SGL-A



Fonte: elaborada pelo autor.

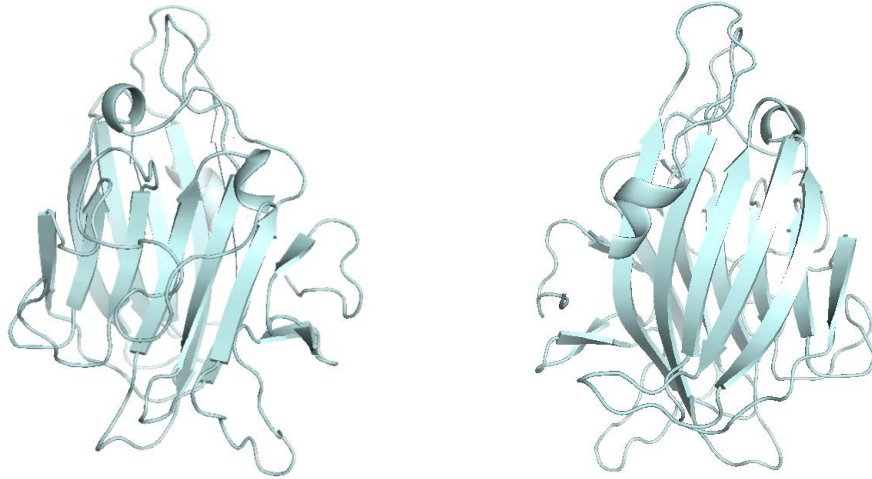
### SGL-E



Fonte: elaborada pelo autor.

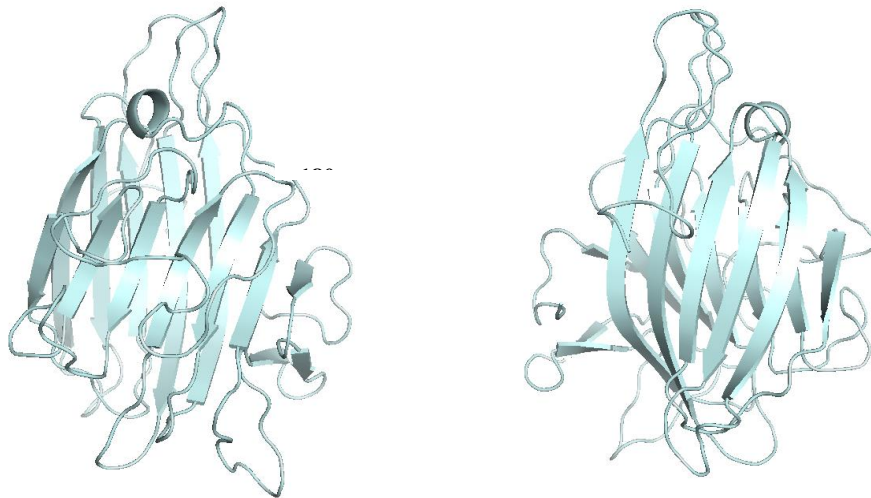
Figura 21 – Modelagem tridimensional das sequências deduzidas da SGL-A e SGL-E, pelo programa M4T

**SGL-A**



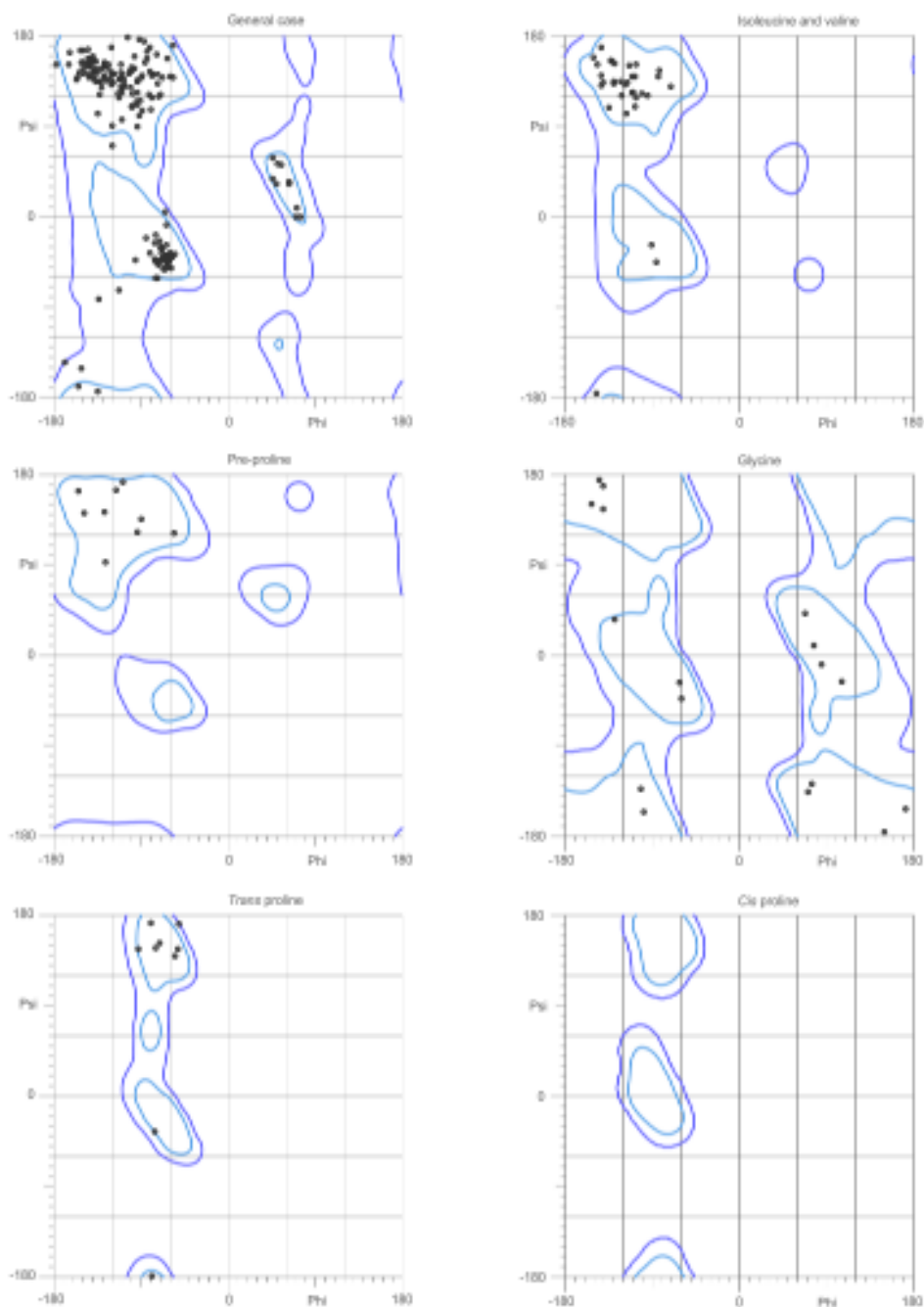
Fonte: elaborada pelo autor.

**SGL-E**



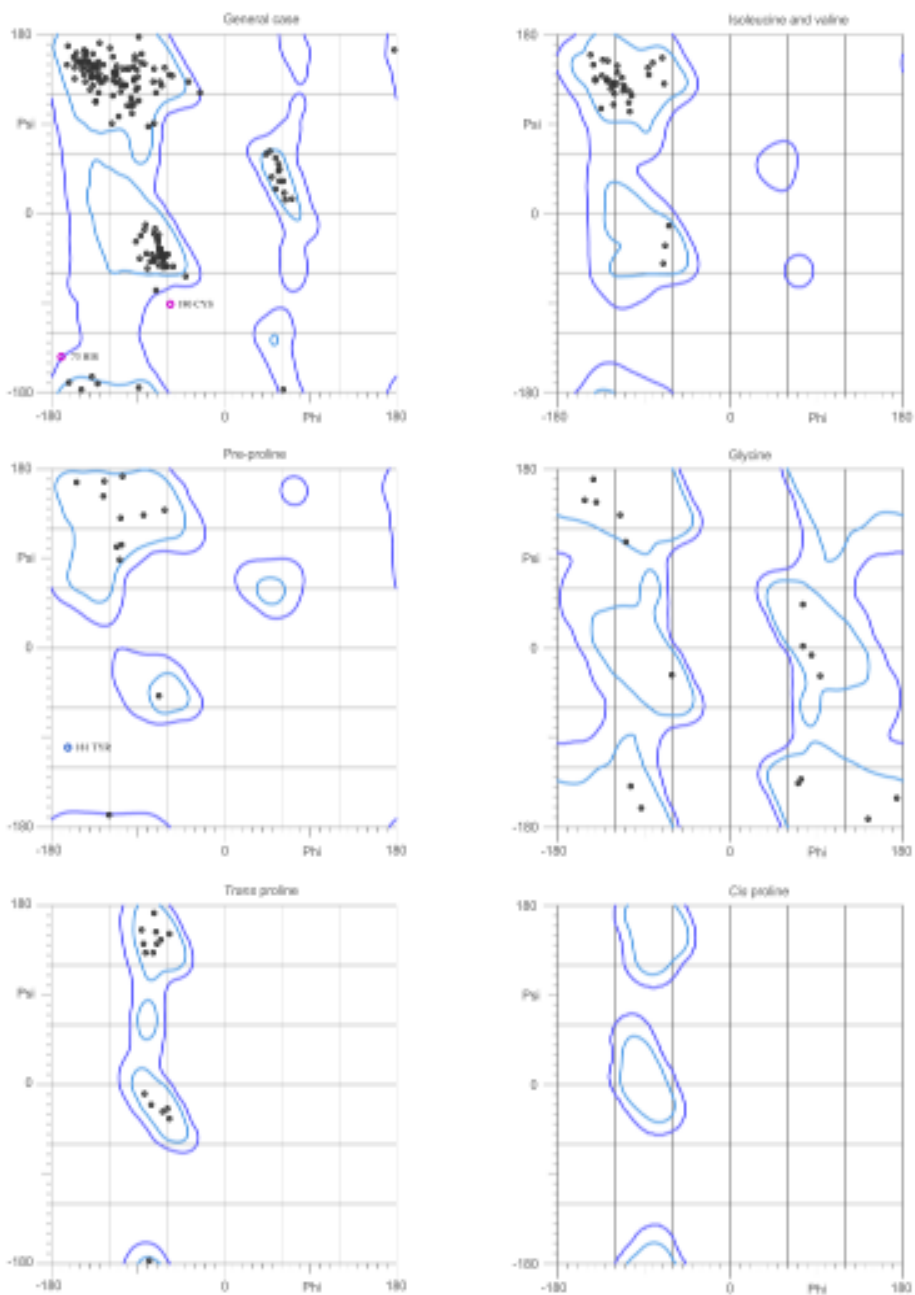
Fonte: elaborada pelo autor.

Figura 22 – Gráfico de Ramachandran, gerado pelo programa MolProbity, para validação da modelagem tridimensional da SGL-A



Fonte: elaborada pelo autor.

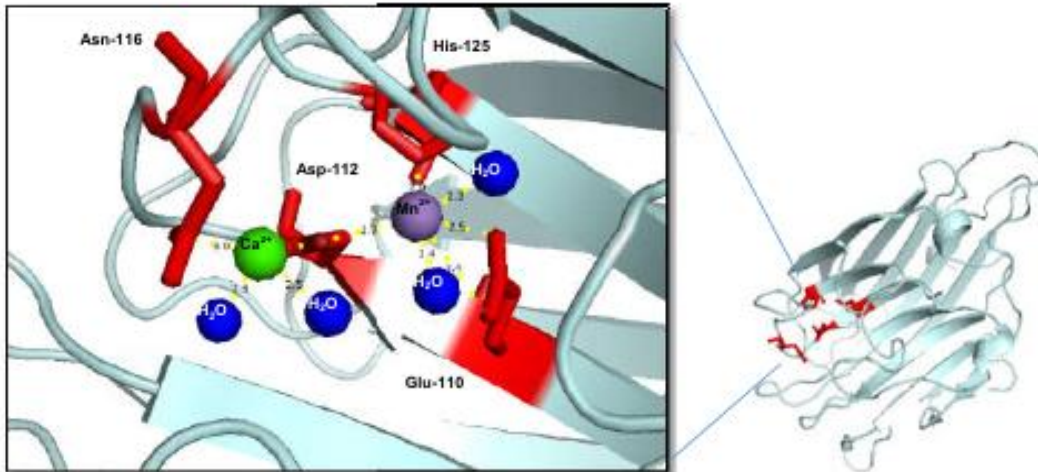
Figura 23 – Gráfico de Ramachandran, gerado pelo programa MolProbity, para validação da modelagem tridimensional da SGL-E



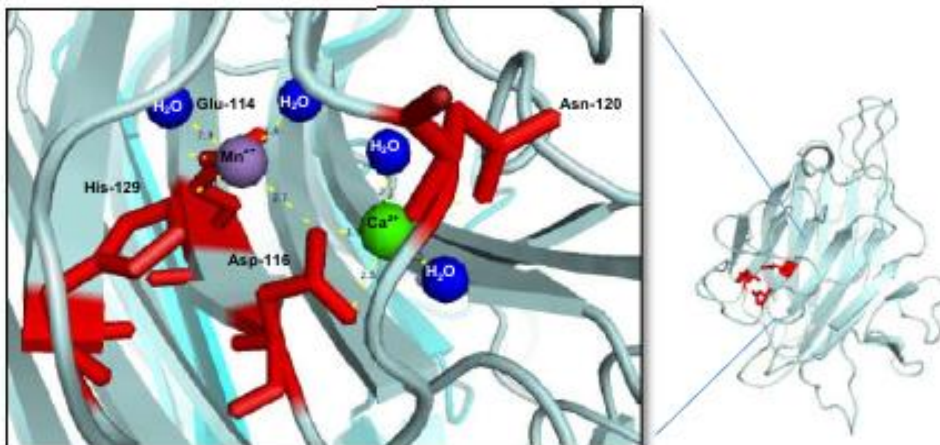
Fonte: elaborada pelo autor.

Figura 24 – Resíduos de aminoácidos da SGL-A e SGL-E, possivelmente, envolvidos na interação com  $\text{Ca}^{2+}$  e  $\text{Mn}^{2+}$ , pelo alinhamento com outras lectinas de leguminosas presentes no NCBI

### SGL-A

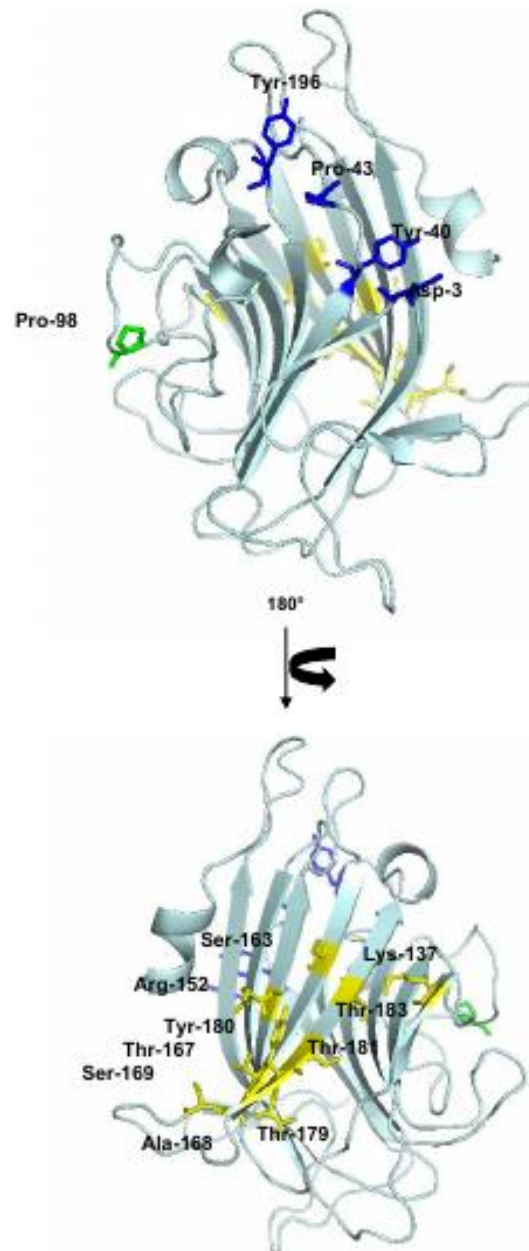


### SGL-E



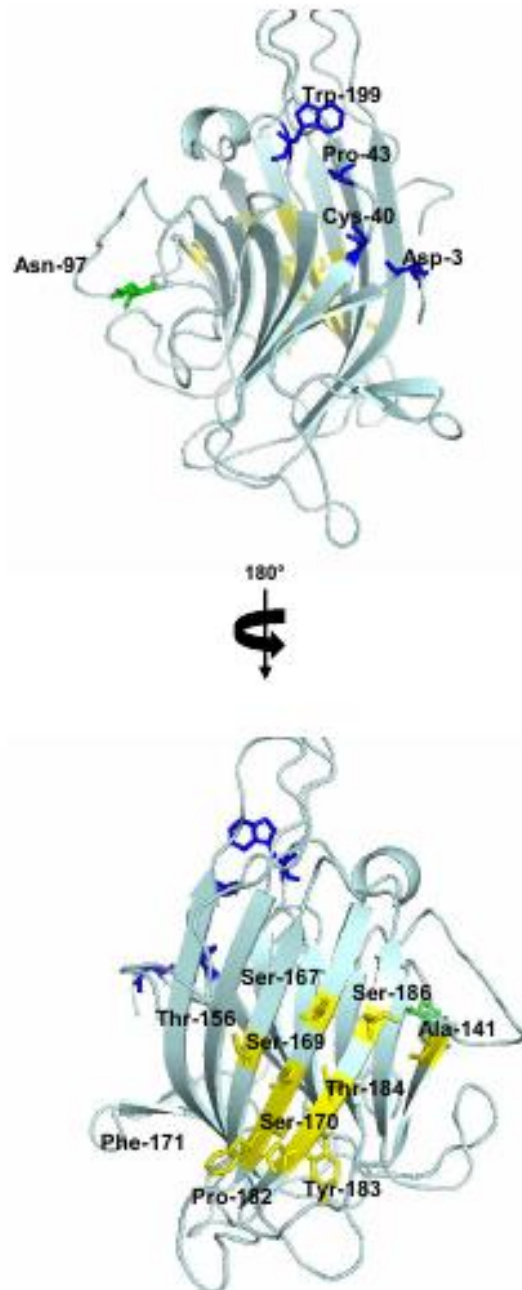
Fonte: elaborada pelo autor.

Figura 25 – Sítios de interações preditos pelo CDD para a SGL-A. O resíduo de cor verde representa o sítio de N-glicosilação, os resíduos de cor azul representam os sítios de interação homodimérica e os resíduos de cor amarela representam os sítios de interações homotetraméricas



Fonte: elaborada pelo autor.

Figura 26 – Sítios de interações preditos pelo CDD para a SGL-E. O resíduo de cor verde representa o sítio de N-glicosilação, os resíduos de cor azul representam os sítios de interação homodimérica e os resíduos de cor amarela representam os sítios de interações homotetraméricas



Fonte: elaborada pelo autor.



O sítio de interação homodimérica da SGL-A é formado por Asp-3, Tyr-40, Pro-43 e Tyr-196. Já o sítio de interação homotetramérica desta mesma sequência é composto por Lys-137, Arg-152, Ser-163, Thr-167, Ala-168, Ser-169, Thr-179, Tyr-180, Thr-181 e Thr-183.

Os resíduos de aminoácidos envolvidos nos sítios de interações homodiméricas da SGL-E são Asp-3, Cys-40, Pro-43 e Trp-199. Quanto aos sítios de interações homotetraméricas da SGL-E, verificamos que são compostos por Ala-141, Thr-156, Ser-167, Ser-169, Ser-170, Phe-171, Pro-182, Tyr-183, Thr-184 e Ser-186.

O índice de hidrofobicidade gerado para as SGL-A e SGL-E foram, respectivamente, -0,36 e -0,18. Estes resultados indicam que as lectinas codificadas por essas sequências são hidrofóbicas. Análises realizadas pelo programa Peptide 2.0 indicam que a lectina deduzida para a SGL-A é composta por 38,3% de resíduos hidrofóbicos, 11,25% de resíduos ácidos, 10,83% de resíduos básicos e 39,58% de resíduos neutros. Essas mesmas análises, quando realizadas para a lectina deduzida para a SGL-E, demonstraram a presença de 40,98% de resíduos hidrofóbicos, 9,84% de resíduos ácidos, 9,43% de resíduos básicos e 39,75% de resíduos de carga neutra.

Os resíduos da SGL-A preditos para fosforilação, obtidos pelo programa NetPhos 2.0 Server, são: Ser-26, Ser-89, Ser-100, Ser-122, Ser-135, Ser-143, Ser-163, Ser-165, Ser-216, Tyr-11, Tyr-114, Tyr-123, Tyr-180, Tyr-184, Tyr-196, Tyr-235, Tyr-239, Thr-21, Thr-58, Thr-83, Thr-138 e Thr-139. Já os resíduos preditos na fosforilação da SGL-E são: Ser-27, Ser-56, Ser-139, Ser-161, Ser-163, Ser-167, Ser-192, Ser-221, Thr-16, Thr-66, Thr-84, Thr-85, Thr-143, Thr-225, Tyr-118 e Tyr-127.

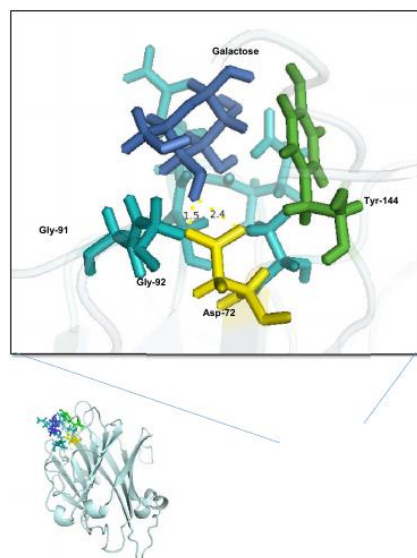
O modelo tridimensional, gerado para a SGL-A, foi escolhido para análise de interação molecular sítio-dirigidas com carboidratos, pois foi o único que demonstrou a presença do resíduo Asp-72 (um resíduo conservado em lectinas de leguminosas e de fundamental importância para o CRD), quando alinhado com outras lectinas de *Fabaceae*. Outros resíduos de aminoácidos apresentaram indícios de que compõem o sítio de reconhecimento a carboidrato da SGL-A, como por exemplo: Ala-71, Gly-91, Gly-98, Tyr-114, Gly-206, Gln-207 e Glu-208.

O docking molecular apresentou uma interação energeticamente favorável com galactose (MDS = -76,24), sendo escolhido, já que este é o carboidrato com o qual a lectina de *Lablabpurpureus* apresenta afinidade. A imagem da interação dos resíduos que formam o CRD da SGL-A, evidenciando o Asp-72 e a Tyr-144 (possível sítio hidrofóbico), estão representados na Figura 26. A interação da SGL-A com N-acetil-D-galactosamina, escolhido para a análise por ser o carboidrato que interage com a lectina de *Vatairea macrocarpa*, foi

predita como sendo energeticamente desfavorável. Esta interação encontra-se ilustrada na Figura 27, na qual podemos perceber que a molécula em questão não chega a adentrar na região composta pelos resíduos de ligação a carboidratos. As distâncias entre os átomos que integram os resíduos de aminoácidos da SGL-A e SGL-E que interagem com os carboidratos, já citados, foram indicadas pelo *software* ProFunc (Figura 28 e 29).

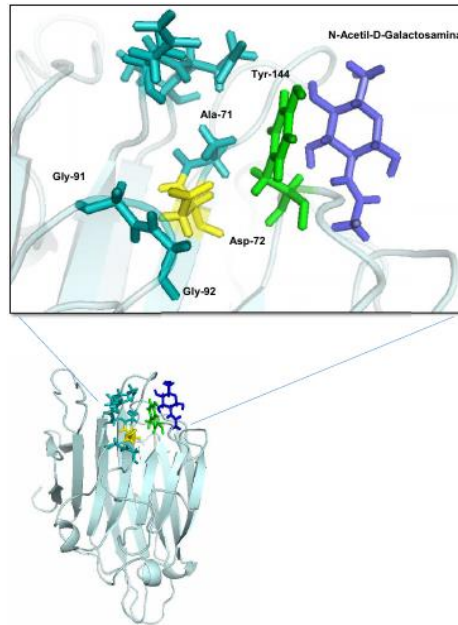
O resultado da Figura 28 mostra que a Ala-71, Gly-92, Tyr-114 e Gln-206 atuam como estabilizadores da ligação da SGL-A com a galactose. Enquanto o NO2 da Asn-116, o OD2 da Asp-72, o OE1 da Gln-208 e o OE1 da Gln-207 realizam uma interação direta com o monossacarídeo. Já na Figura 29, o O2 da N-acetil-D-galactosamina interage com o O da Val-113 e com o O da Pro-70. O O1 do mesmo monossacarídeo interage com o ND2 da Asn-69, enquanto que com o O da Asn-68 não é visto interação com o Asp-72. A Tyr-114, Val-115, Tyr-123 e Arg-124 realizam interações fracas com parte dos átomos da N-acetil-D-galactosamina.

Figura 27 – Interação entre a Galactose e o Domínio de Reconhecimento a Carboidrato da SGL-A, obtida por docking molecular



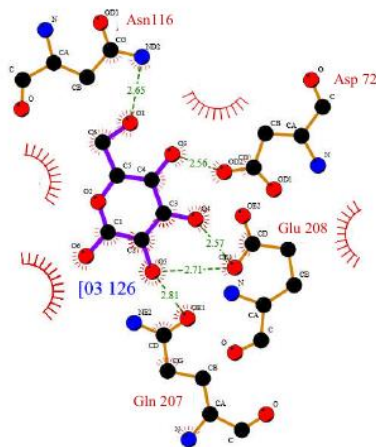
Fonte: elaborada pelo autor.

Figura 28 – Interação entre o N-Acetil-D-Galactosamina e o Domínio de Reconhecimento a Carboidrato da SGL-A, obtida por docking molecular



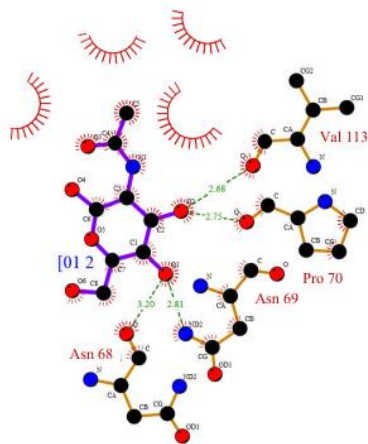
Fonte: elaborada pelo autor.

Figura 29 – Distâncias entre os resíduos de aminoácidos da SGL-A que realizam a interação com a galactose, obtida pelo *software* ProFunc



Fonte: elaborada pelo autor.

Figura 30 – Distância entre os resíduos de aminoácidos da SGL-E que realizam a interação com a galactose, obtida pelo *software* ProFunc



Fonte: elaborada pelo autor.

## 8 DISCUSSÃO

O ensaio de atividade hemaglutinante em extratos proteicos de sementes da família *Fabaceae*, realizado para detecção de lectinas, é pouco eficiente para espécies que possuem alto conteúdo de metabólitos secundários, como por exemplo, espécies do gênero *Swartzia*, pois causam o processo de hemólise dos eritrócitos. Porém, com as sequências nucleotídicas e o uso da bioinformática, o estudo dos genes de novas lectinas pode acrescentar dados substanciais ao processo de prospecção dessa classe proteica. Neste sentido, a presente pesquisa científica trata da caracterização *in silico* de sequências de cDNA, obtida das sementes jovens de *Swartzia grandiflora*, que codificam lectinas. Esta caracterização é de fundamental importância para a compreensão estrutural das lectinas da subtribo Swartzieae, além de contribuir para o entendimento das variações que ocorrem para as lectinas deste grupo taxonômico.

Para tanto, o RNA total do material pesquisado foi, inicialmente, extraído com cloreto de lítio 10M. Este processo de extração rendeu grandes quantidades de RNA, porém não apresentou um alto grau de pureza e integridade. A extração de RNA vegetal de alta qualidade e quantidade é um fator limitante em experimentos de biologia molecular, tais como a amplificação por PCR e síntese de cDNA. Algumas das causas das extrações de RNA tornarem-se pouco eficientes estão relacionadas com as altas concentrações de polissacarídeos e compostos fenólicos nas amostras, os quais podem se co-precipitar ou se ligar ao RNA. Desta forma, podemos especular que a qualidade da amostra de RNA obtido das sementes de *S. grandiflora* tenha sido um dos possíveis motivos para que a amplificação da região 5'UTR não tenha ocorrido de forma eficiente.

O sequenciamento dos resíduos que compõem a região N-terminal da lectina da espécie em estudo foi realizado com o intuito de suprir as informações que seriam obtidas pela RACE 5'. Neste sentido, foi utilizada a banda de 30kDa obtida da eletroforese em gel de poli-acrilamida da farinha das sementes de *S. grandiflora*. A referida banda foi escolhida para sequenciamento por apresentar uma massa molecular característica das lectinas da família *Fabaceae*, além de ser a mesma massa visualizada para a SLL, quando purificada por cromatografia de afinidade à lactose.

Os resíduos de aminoácidos obtidos por sequenciamento apresentaram similaridade, quando submetidos ao banco de sequências de proteínas não-redundantes, como mostramos nas sequências vistas na Tabela 5. A sequência que apresentou maior similaridade aos resíduos obtidos foi com a SJAbg, uma lectina ligante a N-acetil-D-galactosamina, isolada

das raízes de *Sophora japonica*. Além desta lectina, os resíduos apresentaram similaridade com lectinas de *Cladrastis kentukea* e *Robinia pseudoacacia*. Tal resultado demonstra que, provavelmente, grande parte das proteínas presentes na banda de 30kDa, utilizada, são lectinas.

Apenas dez das sequências obtidas pela amplificação da região 3'UTR do cDNA das sementes de *S. grandiflora* apresentaram o domínio conservado das lectinas de leguminosas (L-type). Este compreende o maior domínio de lectinas e o mais bem caracterizado, dentre outros, como o do tipo C- e tipo S-. A existência destes domínios evidencia que este grupo de proteínas é, evolutivamente, conservado na natureza. Este fato pode ser demonstrado quando foram alinhados cerca de 50 lectinas da família *Fabaceae*, sendo visto uma identidade maior de 35 % entre todos os alinhamentos.

Após a tradução das sequências nucleotídicas não redundantes para aminoácidos, foi visualizado, para as seis sequências resultantes, uma elevada similaridade entre seus resíduos (Figura 16). Os pesos moleculares estimados para cada uma dessas sequências são próximos dos valores encontrados para a SLL, na qual se inferiu uma massa molecular de 26000Da estimada por eletroforese em gel de poliacrilamida e que, posteriormente, foi confirmada como sendo de 29000Da por espectrometria de massas. Entretanto, os valores do *pI* deduzidos para as sequências em estudo, contrastaram com o da SLL, mas apresentaram valores aproximados com as lectinas de *Luetzelburgia auriculata* (5,8) e *Vatairea macrocarpa* (4,95).

As sequências deduzidas de aminoácidos da SGL-A e SGL-E, representativas de cada um dos dois subgrupos evidenciados na Tabela 8, foram submetidas a alinhamento pelo NCBI. A maior similaridade vista para as sequências em estudo, utilizando o banco de proteínas não-redundantes, foi com a CLAI (51 %). Isto sugere que, por mais que sejam conservadas, as lectinas de *S. Grandiflora*, identificadas no presente trabalho, constituem um grupo distinto dentre as lectinas de leguminosas caracterizadas até o momento. Adicionalmente, a respeito desta lectina que apresentou homologia de ~51%, é descrito que a CLAI encontrada na espécie *Cladrastis kentukea* apresenta outras lectinas, a saber: CLAI e CLLRP, distinguidas entre si por processamentos pós-traducionais. Algumas das outras lectinas que apresentaram maior similaridade com a SGL-A, nas condições já citadas, são pertencentes à tribo *Sophoreae*, uma tribo filogeneticamente próxima à tribo *Swartzieae*. Dentre essas espécies de *Sophoreae*, encontra-se a Lectina de *Sophora alopecuroides* (SAL). Esta lectina é ligante a D-galactose e apresenta atividade hemaglutinante com eritrócitos humanos e de coelho. Adicionalmente, pode-se dizer que a SAL apresenta atividade

antifúngica e antitumoral. Desta forma, podemos predizer que, possivelmente, a lectina deste estudo possa apresentar alguma aplicação biotecnológica.

As lectinas de *Phaseolus vulgaris* e *Sesbania bisponia* foram algumas das que também apresentaram similaridade com a SGL-E, utilizando o banco de proteínas não-redundantes (Tabela 10). As similaridades vistas para essas lectinas foram, 45 e 46 %, respectivamente. Em relação à PHA (*phytohemagglutinin*), temos que esta é uma lectina tetramérica composta por cinco isoformas dos polipeptídeos PHA-E e PHA-L, em diferentes combinações. Tetrâmeros de PHA-E (34kDa) aglutinam eritrócitos, enquanto que tetrâmeros de PHA-L (32kDa), além de aglutinarem leucócitos, também possuem atividade mitogênica.

As estruturas secundárias, visualizadas na Figura 19, já eram esperadas para as sequências deduzidas de aminoácidos da SGL-A e SGL-E, visto que é conhecido na literatura científica para os monômeros de lectinas de leguminosas, a presença de duas folhas  $\beta$  antiparalelas, uma formada, geralmente, por 7 fitas e a outra, por 5, assim como foi visto neste estudo, ao analisar os modelos tridimensionais gerados pelo M4T server 3.0. A mesma arquitetura e topologia é encontrada em uma grande variedade de proteínas que reconhecem carboidratos, como as galectinas e os “lectin-likedomains Wing-1 e Wing-2”. A topologia estrutural das lectinas em estudo, demonstrada na Figura 19, é complexa e, estruturalmente, relatada como “jelly-roll”. Para evidenciar esta característica nas estruturas secundárias nas lectinas de leguminosas, podemos citar a lectina de *Luetzelburgia auriculata*, uma espécie também do grupo das *Papilionoideae*, encontrada no norte do Brasil, que apresenta 38% de sua estrutura na forma de folhas  $\beta$  e apenas 5% como  $\alpha$ -hélices .

O banco de dados do CDD indicou que os resíduos responsáveis pela interação com  $\text{Ca}^{2+}$  e  $\text{Mn}^{2+}$  da SGL-A e SGL-E também são conservados nas suas estruturas primárias. As únicas exceções conhecidas a esta regra nas lectinas dessa família vegetal é a arcelina e o inibidor de  $\alpha$ -amilase ( $\alpha$ -AI), ambas relatadas para *Phaseolus vulgaris*. Na família das lectinas de *Fabaceae*, já se sabe que os resíduos do sítio de interação a carboidratos adotam uma conformação correta apenas quando na presença dos íons acima citados. Estas interações a metais de transição foram, inicialmente, descritas detalhadamente, para a Con A, na qual é vista a ligação com quatro resíduos de aminoácido e duas ligações com moléculas de água, sendo que uma destas ligações com a água forma uma ponte com o grupo carboxílico do resíduo Asp-208. Entretanto, estes íons, não interagem, diretamente, com o sítio de ligação ao carboidrato. Os resíduos que integram este sítio de interação na SGL-A e SGL-E correspondem, especificamente, aos resíduos Glu-137, Asp-139, Asn-147, Asp-150 e His-155 na PELA e aos resíduos Glu-127, Asp-129 e His-143 da lectinas de *Leucomphalos*

*mildbraedii*.

Lectinas de *Fabaceae* possuem múltiplos sítios de interação e, adicionalmente, alguns desses sítios são conhecidos na formação de dímeros e tetrâmeros. A oligomerização de todas as lectinas envolve, primariamente, interações das seis fitas que formam uma folha  $\beta$ . Na Con A e outras lectinas de *Fabaceae* analisadas subsequentemente, a formação da estrutura dimérica envolve o arranjo lado-a-lado dos dois monômeros, de tal maneira que as duas estruturas formam uma folha  $\beta$  contígua, contendo 12 fitas. Esta estrutura é descrita na literatura como o modo canônico de dimerização das lectinas de *Fabaceae*. Assim sendo, podemos afirmar que esta classe de lectinas formam uma interessante família de proteínas, na qual pequenas alterações na sua estrutura primária geram variações na estrutura terciária, ocasionando, assim, largas variações na estrutura quaternária. Os resíduos averiguados pelo CDD na formação das estruturas homodiméricas e homotetraméricas estão representados nas Figuras 24 e 25. Os resíduos de aminoácidos que compõem esses sítios na SGL-A e SGL-E, demonstram homologia com os resíduos que formam as mesmas regiões em outras lectinas.

O sítio de interação a carboidratos ou CRD das lectinas da família *Fabaceae* são formados por quatro *loops* (A, B, C e D). Estes são compostos por um conjunto de resíduos de aminoácidos, fortemente, conservados. Os resíduos invariantes são a Asp no *loop*A, Gly no *loop* B, Asn e um resíduo aromático (Phe ou Tyr) no *loop* C interagem com o ligante independentemente da configuração do átomo C4. De estudos mutacionais, três resíduos foram identificados como sendo essenciais na ligação ao carboidrato (independentemente de qual seja), a saber: Asp, Asn e um resíduo aromático. Dentre os quatro *loops* envolvidos no reconhecimento a carboidrato, o *loop* D é o mais variável, em termos de tamanho e sequência. Geralmente, um ou dois resíduos de aminoácidos desse *loop* interagem com o ligante, mas os resíduos envolvidos nessa ligação variam de lectina para lectina. O resíduo Asp-72 da SGL-A foi visto por alinhamento estrutural, quando alinhado com outras lectinas de *Fabaceae*. Dentre as lectinas alinhadas, encontra-se a PELA que possui o seu CRD composto por Asp-95, Ala-113, Gly-115, Asn-145, Ser-146, Asn-147, Glu-230 e Gln-231. A ConA, uma lectina ligante a manose assim como a PELA. Já para a SGL-E foi encontrada uma Ala-73 na posição equivalente ao resíduo fundamental (Asp) para a ligação ao carboidrato das outras lectinas.

O docking sítio-dirigido da SGL-A realizado com D-galactose, como já mencionado nos resultados, apresentou uma energia favorável. Já o mesmo procedimento, quando realizado com N-acetil-D-galactosamina, não mostrou indícios para uma possível afinidade com esse carboidrato. Todas as lectinas de *Fabaceae* que reconhecem a galactose ou o N-acetil-D-galactosamina no seu sítio de ligação tem, caracteristicamente, um longo *loop* D,



comparado com as lectinas ligantes a manose, glicose ou fucose. Esse *loop* D na SGL-A é correspondente a Gly-203, Ala-204, Thr-205, Gly-206, Gln-207, Glu-208, Val-209, Ala-210 e Ile-211, sendo diferente do encontrado na PELA.

As estruturas cristalográficas de quatro lectinas ligantes a Gal/GalNac são conhecidas, a saber: as lectinas de *Erythrina corallodendron*, *Arachis hypogea*, *Grifonia simplicifolia* e *Glycine max*. Em todas estas estruturas, o ácido aspártico, conservado no *loop* A dessas lectinas, reconhece o C3-OH equatorial e o C4-OH axial da galactose. O C3-OH faz pontes de hidrogênio adicionais com o NH da Gly no *loop* B e com o Asn no *loop* C, enquanto que C4-OH faz outra ponte de hidrogênio com o *loop* D.

A afinidade relativa dessas lectinas a carboidratos com galactose e galactosamina pode ser atribuída à variação do *loop* C. Um *loop* C longo e contendo resíduos com cadeia lateral grande, geralmente, limita grandes substituições para a galactose nesta posição. A análise das sequências do *loop* C revela que o seu tamanho influencia, significativamente, a especificidade da ligação, como por exemplo, no caso das lectinas com um pequeno *loop* C que interagem com GalNac melhor do que com Gal e lectinas com um grande *loop* C que interagem melhor com Gal do que com GalNac. Adicionalmente, a presença de resíduos com cadeia lateral curta, como Asp, Gly ou Ala, nesse *loop* nas lectinas ligantes a GalNac parece facilitar a interação com o açúcar. O *loop* A da SGL-A é idêntico ao da PELA, sendo formado por Asn-69, Pro-70, Ala-71, Asp-72, Gly-73 e Ile-74. O *loop* B da PELA é formado por Ala, Gln e Gln. Esse *loop* na SGL-A é formado por Gly-90, Gln-91 e Gln-92. Os *loops* C e D da PELA e SGL-A são diferentes, sendo mais um indício de que a SGL-A não seja ligante a manose por não possuir uma composição de aminoácidos que a mesma.

## 9 CONCLUSÃO

As análises das sequências de aminoácidos deduzidas de SGL-A e SGL-E comprovam a presença de lectinas em sementes de *Swartzia grandiflora*, possuindo um domínio de reconhecimento a carboidrato que interage de forma, energeticamente, favorável com galactose. Pode-se concluir, também, que as lectinas de *S. Grandiflora*, identificadas no presente trabalho, constituem um grupo distinto dentre as lectinas de leguminosas caracterizadas até o momento, tendo em vista a identidade observada nos alinhamentos das sequências em estudo.

## REFERÊNCIAS

- ALTSCHUL, S. F. *et al.* Basic local alignment search tool. **Journal of Molecular Biology**. [S. I.]. v. 215. n. 3. p. 403–10. Out. 1990.
- BAKER, R. L. *et al.* A maize trypsin inhibitor (ZmTIp) with limited activity against *Aspergillus flavus*. **Journal of food protection**. [S. I.]. v. 72. n. 1. p. 185–188. 2009.
- BENEVIDES, R. G. *et al.* A lectin from *Platypodium elegans* with unusual specificity and affinity for asymmetric complex N-glycans. **The Journal of biological chemistry**. [S. I.]. v. 287. n. 31. p. 26352–64. 27. Jul. 2012.
- BROOKSBANK, C. *et al.* The European Bioinformatics Institute's data resources. **Nucleic Acids Research**. [S. I.]. v. 31. n. 1. p. 43–50. Jan. 2003.
- BRUNEAU, A. *et al.* **Legume phylogeny and classification in the 21st century: progress, prospects and lessons for other species-rich clades**. v. 62. p. 217–248. April, 2013.
- CALVETE, J. J. *et al.* Amino acid sequence, glycan structure, and proteolytic processing of the lectin of *Vatairea macrocarpa* seeds. **FEBS Letters**. [S. I.]. v. 425. n. 2. p. 286–292. 1998.
- CARDOSO, D. *et al.* Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. **American Journal of Botany**. [S. I.]. v. 99. n. 12. p. 1991–2013. 2012.
- CARDOSO, D. *et al.* Reconstructing the deep-branching relationships of the papilionoid legumes. **South African Journal of Botany**. [S. I.]. v. 89. p. 58–75. 2013.
- CARLINI, C. R.; GROSSI-DE-SÁ, M. F. Plant toxic proteins with insecticidal properties. A review on their potentialities as bioinsecticides. **Toxicon**. [S. I.]. v. 40. n. 11. p. 1515–1539. 2002.
- CARVALHO, A. D. S. *et al.* Purification, characterization and antibacterial potential of a lectin isolated from *Apuleia leiocarpa* seeds. **International Journal of Biological Macromolecules**. [S. I.]. v. 75. p. 402–408. 2015.
- CAVALCANTI, M. S. M.; COELHO, L. C. B. B. Isolation and partial purification of lectin from *Swartzia pickellii* Killip (White Jacaranda). **Mem. Inst. Oswaldo Cruz**. [S. I.]. v. 85. n. 3. p. 371–372. 1990.
- CHANG, S.; PURYEAR, J.; CAIRNEY, J. A simple and efficient method for isolating RNA from pine trees. **Plant Molecular Biology Reporter**. [S. I.]. v. 11. n. 2. p. 113–116. 1993.
- CHEN, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. **Acta Crystallographica Section D: Biological Crystallography**. [S. I.]. v. 66. n. 1. p. 12–21. 2010.
- CHRISPEELS, M. J.; RAIKHEL, N. V. Lectins, lectin genes, and their role in plant defense. **The Plant cell**. [S. I.]. v. 3. n. 1. p. 1–9. 1991.

- COUTINHO, L. M. O conceito de bioma. **Acta Botanica Brasilica**. [S.I.]. v. 20. n. 1. p. 13–23. 2006.
- DAMME, E. J. M. VAN *et al.* In the bark of yellow wood (*Cladrastis lutea*). p. 579–598. 1995.
- DE SOUZA CANDIDO, E. *et al.* Plant storage proteins with antimicrobial activity: novel insights into plant defense mechanisms. **The FASEB Journal**. [S.I.]. v. 25. n. 10. p. 3290–3305. 2011.
- EWING, B. *et al.* Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Research**. [S.I.]. v. 8. n. 3. p. 175–185. Mar. 1998.
- EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**. [S.I.]. v. 8. n. 3. p. 186–194. Mar. 1998.
- FERNANDES, A. V. *et al.* Purification and Characterization of a Lectin of the Swartzieae Legume Taxa. **Protien & Peptide Letter**. [S.I.]. v. 19. p. 1082–8. 2012.
- FERNANDES, A. V. *et al.* Lectin genes and their mature proteins: Still an exciting matter, as revealed by biochemistry and bioinformatics analyses of newly reported proteins. **Biochemical Systematics and Ecology**. [S.I.]. v. 60. p. 46–55. 2015.
- FERNANDES, A. V. *et al.* Seeds of Amazonian *Fabaceae* as a source of new lectins. **Brazilian Journal of Plant Physiology**. [S.I.]. v. 23. n. 3. p. 237–244. 2011.
- FERNANDEZ-FUENTES, N. *et al.* M4T: A comparative protein structure modeling server. **Nucleic Acids Research**. [S.I.]. v. 35. n. SUPPL. 2. p. 363–368. 2007a.
- FERNANDEZ-FUENTES, N. *et al.* Comparative protein structure modeling by combining multiple templates and optimizing sequence-to-structure alignments. **Bioinformatics**. [S.I.]. v. 23. n. 19. p. 2558–2565. 2007b.
- FITCHES, E. *et al.* The insecticidal activity of recombinant garlic lectins towards aphids. **Insect Biochemistry and Molecular Biology**. [S.I.]. v. 38. n. 10. p. 905–915. 2008.
- GORDON, D.; ABAJIAN, C.; GREEN, P. Consed: a graphical tool for sequence finishing. **Genome Research**. [S.I.]. v. 8. n. 3. p. 195–202. Mar. 1998.
- HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series**. [S.I.]. v. 41. p. 95–98. 1999.
- HE, S. *et al.* Identification of a lectin protein from black turtle bean (*Phaseolus vulgaris*) using LC-MS/MS and PCR method. **LWT - Food Science and Technology**. [S.I.]. v. 60. n. 2. p. 1074–1079. 2015.
- KELLEY, L. A. *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. **Nature Protocols**. [S.I.]. v. 10. n. 6. p. 845–858. 2015.

- KENNEDY, J. F. *et al.* Lectins, versatile proteins of recognition: a review. **Carbohydrate Polymers**. [S.I.]. v. 26. n. 3. p. 219–230. 1995.
- KOENEN, E. J. M. *et al.* Exploring the tempo of species diversification in legumes. **South African Journal of Botany**. [S.I.]. v. 89. p. 19–30. 2013.
- LAEMMLI, U. K. Technique Electrophorèse. **Nature**. [S.I.]. v. 227. p. 680–685. 1970.
- LAM, S. K.; NG, T. B. Lectins: Production and practical applications. **Applied Microbiology and Biotechnology**. [S.I.]. v. 89. n. 1. p. 45–55. 2011.
- LI, T. *et al.* Isolation and characterization of a novel lectin with antifungal and antiproliferative activities from *Sophora alopecuroides* seeds, p. 606–613. May. 2012.
- LI, Y. *et al.* Molecular cloning, expression, and characterization of a *Sophora alopecuroides* lectin from *Escherichia coli*. v. 46. n. 9. p. 749–756. 2014.
- LORIS, R. *et al.* Legume lectin structure. Biochimica et Biophysica Acta - **Protein Structure and Molecular Enzymology**. [S.I.]. v. 1383. n. 1. p. 9–36. 1998.
- LORIS, R. Principles of structures of animal and plant lectins, Biochim. Biophys. **Acta**. [S.I.]. v. 1572. p. 198–208. 2002.
- MARQUI, S. R. DE *et al.* Saponinas antifúngicas de *Swartzia langsdorffii*. **Química Nova**. [S.I.]. v. 31. n. 4. p. 828–831. 2008.
- MORENO, F. B. M. B. *et al.* Identification of a new quaternary association for legume lectins. **Journal of Structural Biology**. [S.I.]. v. 161. n. 2. p. 133–143. 2008.
- NASCIMENTO, K. S. *et al.* An overview of lectins purification strategies. **Journal of Molecular Recognition**. [S.I.]. v. 25. n. 11. p. 527–541. 2012.
- NASI, A.; PICARIELLO, G.; FERRANTI, P. Proteomic approaches to study structure, functions and toxicity of legume seeds lectins. Perspectives for the assessment of food quality and safety. **Journal of Proteomics**. [S.I.]. v. 72. n. 3. p. 527–538. 2009.
- OLIVEIRA, J. T. A *et al.* Purification and physicochemical characterization of a cotyledonary lectin from *Luetzelburgia auriculata*. **Phytochemistry**. [S.I.]. v. 61. n. 3. p. 301–310, 2002.
- PEUMANS, W. J.; VAN DAMME, E. J. Lectins as plant defense proteins. **Plant physiology**. [S.I.]. v. 109. n. 2. p. 347–352. 1995.
- PINHEIRO, A. Q. *et al.* Antifungal and marker effects of *Talisia esculenta* lectin on *Microsporum canis* *in vitro*. **Journal of Applied Microbiology**. [S.I.]. v. 107. n. 6. p. 2063–2069. 2009.
- RABIJNS, A. *et al.* A legume lectin from the bark of *Robinia pseudoacacia* crystallizes in two crystal forms: Preliminary diffraction analyses. Acta Crystallographica Section D: **Biological Crystallography**. [S.I.]. v. 56. n. 12. p. 1638–1640. 2000.

RAO, V. S.; LAM, K.; QASBA, P. K. Architecture of the sugar binding sites in carbohydrate binding proteins--a computer modeling study. **International journal of biological macromolecules**. [S.I.]. v. 23. n. 4. p. 295–307. 1998.

RIBEIRO, A. *et al.* Are Vicilins Another Major Class of Legume Lectins? **Molecules**. [S.I.]. v. 19. n. 12. p. 20350–20373. 2014.

ROCHA, B. A. M. *et al.* CRLI induces vascular smooth muscle relaxation and suggests a dual mechanism of eNOS activation by legume lectins via muscarinic receptors and shear stress. **Archives of Biochemistry and Biophysics**. [S.I.]. v. 565. p. 32–39. 2015.

RYKUNOV, D. *et al.* Improved scoring function for comparative modeling using the M4T method. **Journal of Structural and Functional Genomics**. [S.I.]. v. 10. n. 1. p. 95–99. 2009.

SÁ, R. A. *et al.* Larvicidal activity of lectins from *Myracrodruon urundeuva* on *Aedes aegypti*. *Comparative Biochemistry and Physiology – C. Toxicology and Pharmacology*. [S.I.]. v. 149. n. 3. p. 300–306. 2009.

SAH, S. K.; KAUR, G.; KAUR, A. Rapid and Reliable Method of High-Quality RNA Extraction from Diverse Plants. **American Journal of Plant Sciences**. [S.I.]. v. 5. n. 5. p. 3129–3139. 2014.

SAMBROOK, J.; FRITSCH, E.; MANIATIS, T. *Molecular Cloning: A Laboratory Manual*. 2nd. ed. Cold Spring Harbor: **Cold Spring Harbor Laboratory Press**. [S.I.]. 1989.

SHARMA, V.; SUROLIA, A. Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. **Journal of molecular biology**. [S.I.]. v. 267. n. 2. p. 433–445. 1997.

SHARON, N.; LIS, H. Legume lectins--a large family of homologous proteins. **The FASEB journal: official publication of the Federation of American Societies for Experimental Biology**. [S.I.]. 1990.

SHARON, N.; LIS, H. History of lectins: From hemagglutinins to biological recognition molecules. **Glycobiology**. [S.I.]. v. 14. n. 11. p. 53–62. 2004.

SHETTY, K. N. *et al.* Affinity of a galactose-specific legume lectin from *Dolichos lablab* to adenine revealed by X-ray crystallography. **IUBMB Life**. [S.I.]. v. 65. n. 7. p. 633–644. 2013.

SINGHA, S. *et al.* Comparison of the nature of interactions of two sialic acid specific lectins *Saraca indica* and *Sambucus nigra* with N-acetylneuraminic acid by spectroscopic techniques. **Journal of Luminescence**. [S.I.]. v. 160. p. 119–127. 2015.

THOMSEN, R.; CHRISTENSEN, M. H. MolDock: a new technique for high accuracy molecular docking. **J. Med. Chem.** [S.I.]. v. 49. p. 3315–3321. 2006.

TORKE, B. M.; SCHAAL, B. A. **Molecular phylogenetics of the species-rich neotropical genus SWARTZIA (Leguminosae, Papilionoideae) and related genera of the swartzoid clade 1**. [S.I.]. v. 95. n. 2. p. 215–228. 2008.

TOWBIN, H.; STAEBELIN, T.; GORDON, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. **Proceedings of the National Academy of Sciences of the United States of America**. [S.I.]. v. 76. n. 9. p. 4350–4354. 1979.

UNITT, J.; HORNIGOLD, D. Plant lectins are novel Toll-like receptor agonists. **Biochemical Pharmacology**. [S.I.]. v. 81. n. 11. p. 1324–1328. 2011.

VAN DAMME, E. J. *et al.* The bark of *Robinia pseudoacacia* contains a complex mixture of lectins. Characterization of the proteins and the cDNA clones. **Plant physiology**. [S.I.]. v. 107. n. 3. p. 833–843. 1995.

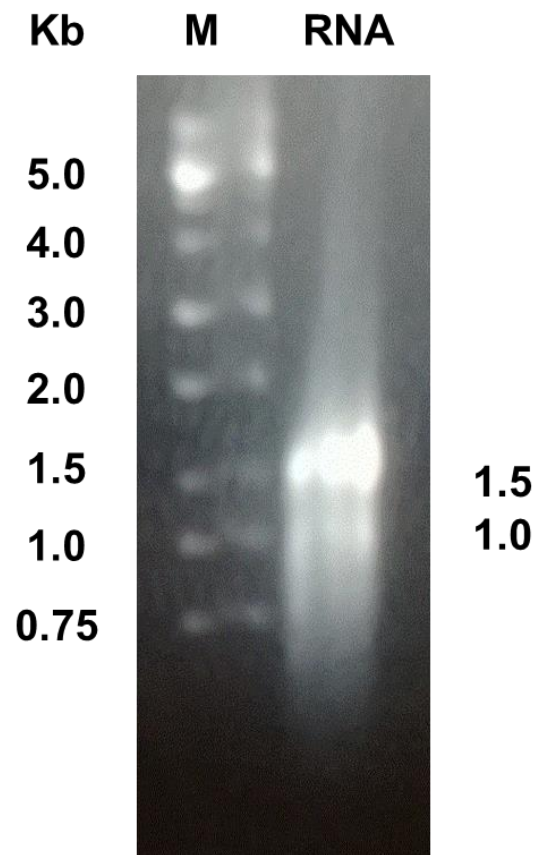
VAN DAMME, E. J. M. *et al.* Molecular cloning of the bark and seed lectins from the Japanese pagoda tree (*Sophora japonica*). **Plant Molecular Biology**. [S.I.]. v. 33. n. 3. p. 523–536. 1997.

VIJAYAN, M.; CHANDRA, N. Lectins. **Current opinion in structural biology**. [S.I.]. v. 9. n. 6. p. 707–714. 1999.

YU, D. *et al.* Comparison and Improvement of Different Methods of RNA Isolation from Strawberry (*Fragria \* ananassa*). **Journal of Agricultural Science**. [S.I.]. v. 4. n. 7. p. 51–56. 2012.

**APÊNDICES****APÊNDICE A – ELETROFORESE EM GEL DE AGAROSE 1,5% DO RNA  
EXTRAÍDO DAS SEMENTES DE *SWARTZIA GRANDIFLORA***

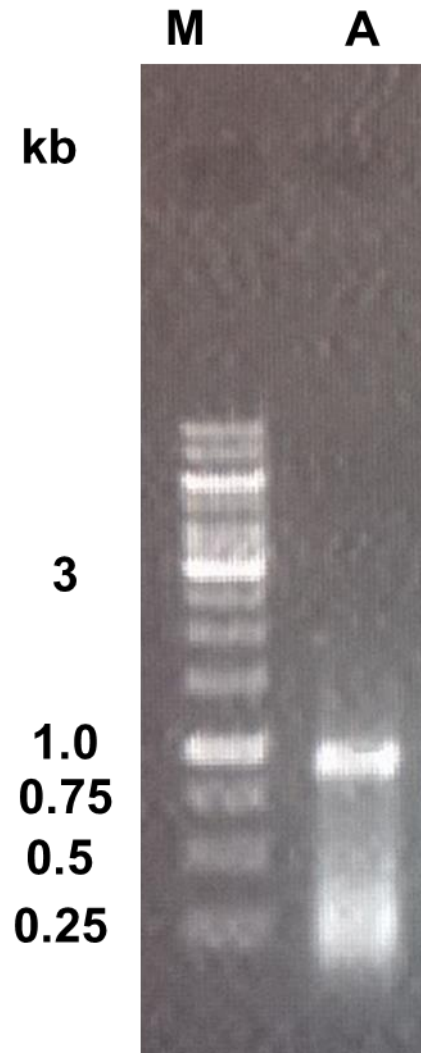
Nota: Foi aplicado, aproximadamente, 1.000ng de RNA total, utilizando-se como marcador (M) o DNA Scada Ladder 1 kb.





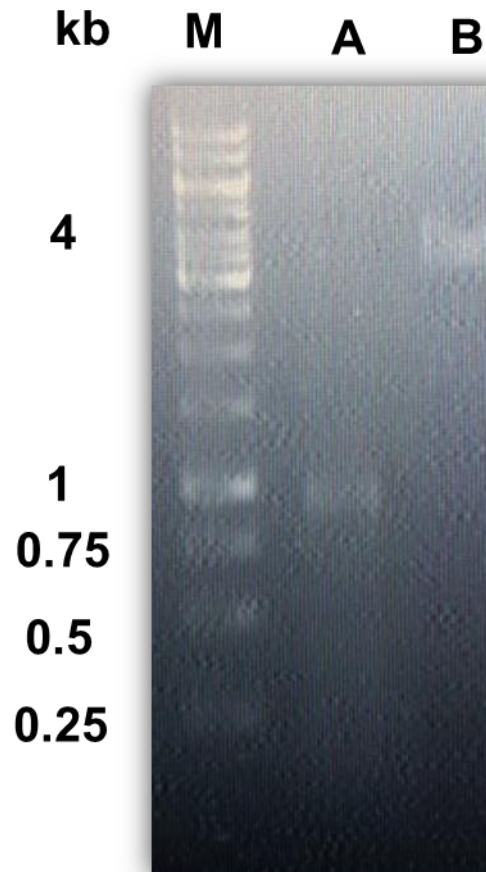
**APÊNDICE B – ELETROFORESE EM GEL DE AGAROSE 1% DO PRODUTO DA  
RACE 3' DO RNA EXTRAÍDO DE *SWARTZIA GRANDIFLORA***

Nota: Foi aplicado, aproximadamente, 1.000ng de RNA. A letra M indica a posição do marcado molecular Scada Ladder 1 kb.



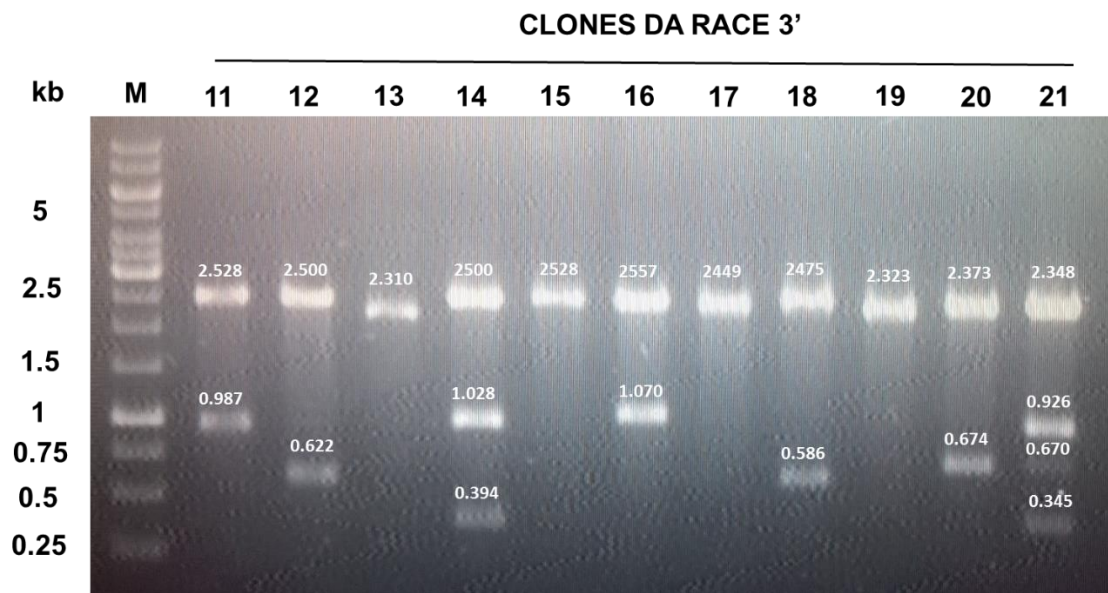
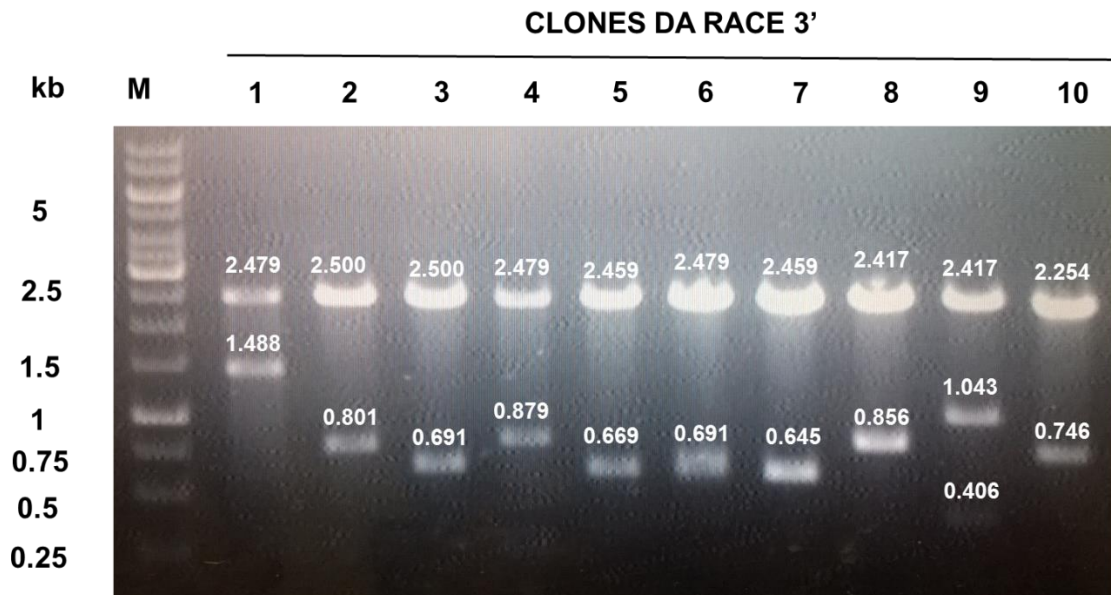
**APÊNDICE C – ELETROFORESE EM GEL DE AGAROSE 1% DO PRODUTO DA  
RACE 3' DO RNA EXTRAÍDO DE *SWARTZIA GRANDIFLORA* (A) E DESSE  
MESMO PRODUTO LIGADO AO PGEM - T(B)**

Nota: A letra M indica o poço, contendo o marcado molecular Scada Ladder 1 kb.



**APÊNDICE D – ELETROFORESE EM GEL DE AGAROSE 1% DO PRODUTO DA DIGESTÃO COM *PVUII* DOS CLONES TRANSFORMADOS COM OS PRODUTOS DA RACE 3' DO RNA EXTRAÍDO DE *SWARTZIA GRANDIFLORA***

Nota: A letra M indica o poço, contendo o marcado molecular Scada Ladder 1 kb.



## APÊNDICE E – FONTES DE AUXÍLIO E FINANCIAMENTO

Esta pesquisa científica foi realizada com o auxílio das seguintes instituições:

- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), através do Programa Reuni, de Orientação e Operacionalização da Pós-Graduação, articulada à Graduação (PROPAG), pela concessão da bolsa de pós-graduação ao autor deste estudo;
- Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), através do Projeto CNPq-Bionorte, N° 554307/2010-3, cujo título é “Moléculas com potencial antifúngico em sementes, folhas e casca de espécies arbóreas: bioprospecção, uso e sustentabilidade da flora amazônica”;
- Universidade Federal do Ceará, Departamento de Bioquímica e Biologia Molecular do Centro de Ciências, que financiou as atividades de coleta e em cujos laboratórios foi realizada esta pesquisa.

## ANEXO

**ANEXO A – MAPA DO VETOR DE CLONAGEM PGEM -T, DESTACANDO OS DIVERSOS SÍTIOS DE RESTRIÇÃO DE DIFERENTES ENZIMAS, A REGIÃO DO GENE QUE CONFERE RESISTÊNCIA À AMPICILINA E O GENE LACZ**

