horticulturae

MDPI

Article

Characterization of Lectin from *Bauhinia holophylla* Using Bioinformatics Tools

Ludmila Maria Gonçalves Godoi de Camargos [1], Débora de Oliveira Lopes [2,*] and Ana Hortência Fonseca Castro [1,*]

[1] Laboratory of Botany and Medicinal Plants, Postgraduate Program in Biotechnology, Campus Centro-Oeste Dona Lindu, Federal University of São João del-Rei (UFSJ), Avenida Sebastião Gonçalves Coelho, 400-Chanadour, Divinópolis 35501-296, MG, Brazil; ludgodoi.camargos@gmail.com

[2] Laboratory of Molecular Biology, Campus Centro-Oeste Dona Lindu, Federal University of São João del-Rei (UFSJ), Avenida Sebastião Gonçalves Coelho, 400-Chanadour, Divinópolis 35501-296, MG, Brazil

* Correspondence: debora@ufsj.edu.br (D.d.O.L.); acastro@ufsj.edu.br (A.H.F.C.)

Abstract: Lectins are proteins of a non-immune nature with activity against microorganisms, insects, and tumor cells. The aim of this work was to predict the physicochemical characteristics, structure, and functional properties of a *Bauhinia holophylla* lectin (BhL), sequenced from genomic material obtained from calli cultures, through bioinformatics tools. The results showed a high similarity between the *Bhl* gene and nucleotide sequences that encode lectins expressed by *Bauhinia* species and a high identity between the protein sequence of BhL and lectins from *B. forficata* (90%), *B. variegata* (79.04%), *B. purpurea* (78.01%), and *B. ungulata* (85.27%). BhL has 289 amino acids, of which 30, 85, and 174 residues are related to α-helix, β-sheet, and disordered regions, respectively. Their estimated molecular weight is 31.9 kDa and the theoretical isoelectric point is 5.79. *Bauhinia holophylla* lectin possibly undergoes phosphorylation and glycosylation at specific sites. Conserved protein domains, catalytic sites, and conserved amino acids were observed in BhL, bringing it closer to lectin families from other legume species. The prediction signaled the presence of a sequence of 28 amino acids at the N-terminal end of BhL, with a high hydropathicity index and conceptualized as a signal peptide. The molecular function predicted for BhL was associated with carbohydrate recognition activity. BhL could be an extracellular protein, and its three-dimensional structure showed 78.82% identity with the *B. purpurea* lectin.

Keywords: computational biology; *Bauhinia* species; Leguminosae; in silico analysis

Citation: Camargos, L.M.G.G.d.; Lopes, D.d.O.; Castro, A.H.F. Characterization of Lectin from *Bauhinia holophylla* Using Bioinformatics Tools. *Horticulturae* **2024**, *10*, 682. https://doi.org/10.3390/horticulturae10070682

Academic Editors: Pascual García-Pérez and Lucía Cassani

Received: 13 May 2024
Revised: 21 June 2024
Accepted: 24 June 2024
Published: 27 June 2024

1. Introduction

Lectins correspond to a heterogeneous group of proteins of a non-immune nature that specifically recognize and bind reversibly to carbohydrate structures but do not possess enzymatic activity [1]. They are soluble or membrane proteins present both inside and outside cells [2]. Calcium and manganese ions are required for its structuring, sugar binding, and biological activities [3,4]. Lectins are widely expressed in diverse organisms, including fungi, bacteria, viruses, and higher plants, and are also known as agglutinins or phytohemagglutinin due to their ability to agglutinate red blood cells and/or precipitate glycoconjugates [4,5]. Generally, lectins are encoded by gene families, with highly homologous sequences that differ in some amino acids, resulting in different isoforms [1]. In the plant kingdom, most known lectins have been isolated from species of the Leguminosae family, which includes the *Bauhinia* genus. Taxonomically, Leguminosae has three subfamilies, namely, Caesalpinioideae, Mimosoideae, and Papilionoideae, the latter of which most of the studies on lectins concentrate on [3,6]. *Bauhinia* species are popularly known as "pata-de-vaca" and comprise more than 300 species of the Leguminosae family, Caesalpinioideae subfamily, of tropical occurrence; they are traditionally used in folk medicine to treat infections, pain, and diabetes [7,8].

Generally, legume lectins present high similarities in primary, secondary, and tertiary structures and exhibit comparable physicochemical and structural properties. However, lectins differ considerably in relation to their quaternary structure and specificity to carbohydrates; therefore, they can be involved in various biological processes [4–6]. Plant lectins are fundamental to plant life and have important roles in cell-to-cell communication, stress signal transduction, development, and defense strategies [1,2,9–11]. In addition, the therapeutic potential of lectins with anti-inflammatory [12,13], antinociceptive [13], antitumor [14,15], and antimicrobial effects [16] has been demonstrated.

Obtaining proteins directly from plants is still a challenge since this process generally involves an expensive and low-yield process. Furthermore, the functional characterization and detailed study of many lectins can be compromised by the quantity and purity of the protein obtained through processes traditionally used in plant extraction [17,18]. Generally, the presence of different lectin isoforms in plant extracts prevents a more detailed study of these heterogeneous mixtures, which makes it difficult to conduct other studies to evaluate the functional properties and biological activities of a specific lectin [19].

Advances observed in recent decades in the molecular biology and biotechnology fields have allowed genome sequencing of several species and expanded genomic and proteomic data available in public repositories. There are over 50,000 genomes from eukaryotes, viruses, plasmids, and organelles deposited in the National Center for Biotechnology Information (NCBI) genome database (https://www.ncbi.nlm.nih.gov/genome/, accessed on 25 October 2023) [20]. This large amount of available data has led to the improvement of bioinformatics tools, which are useful for characterizing proteins and their isoforms present in a specific cell or tissue and predicting their functions, providing important information for future in vivo and in vitro studies [21–23].

Currently, in silico functional studies play an important role in predicting protein function, linking it to one or more cellular processes, and helping to elucidate its role in certain metabolic pathways [24]. Bioinformatics has been widely used to provide structural and functional predictions and interactions with other proteins and to allow evolutionary inferences to be made [25]. This tool uses computer algorithms to analyze biological data, including genetic information, protein amino acid sequences, and protein structures [26]. Several programs with friendly and intuitive interfaces are available for free. Using algorithms that mimic the conditions and situations that occur in nature, these platforms allow us to seek answers to complex biological questions, saving time and money [25,27]. In this way, the in silico characterization of lectins not only provides greater depth in biological investigations but also allows an individualized study of proteins of interest and better planning of experimental steps in vitro and in vivo [23].

Despite their importance and the fact that lectins have been studied for a long time, studies on *Bauhinia holophylla* lectins, a native species from the Brazilian Cerrado [6,28], still remain scarce. A single study carried out by Castro et al. [29] demonstrated the possibility of producing in vitro lectins from *B. holophylla* via callus culture. Therefore, the present study aimed to obtain a lectin isoform from *B. holophylla* callus and clone and sequence the gene to characterize the protein using computational tools. The results obtained in this study will facilitate the design and optimization of future studies on the production of recombinant lectins from *B. holophylla* and the proposal of functional tests to validate the in silico analyses.

2. Materials and Methods

2.1. Chemicals

For this study, 6-benzylaminopurine, concanavaline A, trypsine, cetylethylammonium bromide (CTAB), agarose, ethidium bromide, 2x YT medium, ampicillin, and pUC18 plasmid were purchased from Sigma-Aldrich Co. (St. Louis, MO, USA). DNA polymerase (Phusion Hot Start II High-Fidelity DNA Polymerase, Thermo Scientific, Mundelein, IL, USA), Taq DNA polymerase (Ludwig Biotecnologia, Rio Grande do Sul, Brazil), pGEM

T-Easy (Promega, WI, USA), the QIAquick Gel Extraction Kit, and the Plasmid Miniprep Kit were acquired from QIAGEN (North Rhine-Westphalia, Germany).

2.2. Plant Material, Callus Culture, and Hemagglutination Assay

Bauhinia holophylla (Bong.) Steud. (Fabaceae: Cercidoideae) seeds were collected in the Brazilian Cerrado in Ijaci, Southern Minas Gerais State, Brazil (21°09′97″ S and 44°55′65″ W GRW, at 835 m altitude) (SISBIO n° 24542-3, IBAMA Registration: 5042260). Fertile samples were collected, and vouchers were identified by Andreia Fonseca Silva of the PAMG Herbarium (PAMG 57021) at the Agricultural Research Company of Minas Gerais (EPAMIG). This study had access permission for plant genetic heritage components (No. 010500/2014-6/CNPq/CGEN/MMA) and was registered on the SisGen Platform (Register ADFB4A8), according to the Brazilian Biodiversity Law (13.123/2015). This work was also submitted and approved by the Human Research Ethics Committee of the Federal University of São João del-Rei (CAAE: 44407121.4.0000.5545).

Genomic DNA of *B. holophylla* was obtained from 60-day-old callus cultures. Calli were induced in the Wood Plant Medium (WPM) [30] supplemented with 4.44 µM of 6-benzylaminopurine (BAP) in the presence of light at 27 °C with a photoperiod of 16 h, according to Castro et al. [29].

Fresh callus samples were used to confirm the presence of *B. holophylla* lectin (BhL) through the assessment of hemagglutinating activity. For this, lectins were extracted from fresh callus using saline solution (0.9% sodium chloride) [31]. The extract obtained was lyophilized and used to determine the hemagglutinating activity using human blood. A hemagglutination assay was performed using the serial dilution technique of Liener and Hill [32]. The negative control consisted of the calcium tris-buffered saline (CTBS) buffer and the erythrocyte solution, and as a positive control, 0.05% Concanavalin A (obtained from *Canavalia ensiforms*) was used in CTBS buffer (pH = 7.5) [33]. The assays were performed in triplicate.

2.3. Obtaining the Nucleotide Sequence

2.3.1. Primer Construction

The specific primers for the amplification of the *B. holophylla* lectin coding sequence were constructed from the alignment and identification of conserved regions between *B. forficata* (GenBank no. KX027277.1), *B. purpurea* (GeneBank no. D12481.1), and *B. variegata* lectin (GeneBank no. EF558621.1) gene sequences (Table 1) since, in the NLM/NCBI database (GenBank), there was no deposit of annotations referring to the *Bhl* gene.

Table 1. Forward and reverse primers were constructed for the amplification of the lectin-coding gene in *B. holophylla*.

Primers	Sequence (5′→3′)	Tm (°C)
Forward	**TAAGCA**[1]GGATCC [2]ATGCTTCTCTACAACTCAA	69.5
Reverse	**TAAGCA**[1]CTCGAG [3]CATACTGGAATAAGAGGC	69.5

[1] Protective sequence. [2] BamHI restriction site. [3] XhoI restriction site.

2.3.2. Extraction of Genomic DNA

The genomic DNA of *B. holophylla* was extracted from fresh and compact calli grown in vitro for 60 days. A 2% CTAB solution associated with glass microspheres was used for extraction. The integrity and quality of the extracted DNA were assessed using 1% agarose gel electrophoresis and staining with an ethidium bromide solution. The extracted DNA was also quantified using a NanoDrop spectrophotometer (NanoDrop 2000c, Fisher Scientific™, Marietta, GA, USA).

2.3.3. Gene Amplification and Cloning

Amplification of the *Bhl* gene was carried out using the polymerase chain reaction (PCR) with *B. holophylla* genomic DNA and specific primers. The PCR reaction was prepared in a final volume of 20 µL, following the pipetting and volume guidelines specified for the DNA polymerase enzyme. A thermal cycler (Applied Biosystems Veriti^{TM} Dx Thermal Cycler, Thermo Fisher Scientific^{TM}, Marietta, GA, USA) was used, and the amplification program (especially the annealing temperature) was adjusted through previous temperature gradient tests. The optimized program for the amplification of the *Bhl* gene consisted of 3 min of initial denaturation at 98 °C, 35 cycles of denaturation at 95 °C for 10 s, annealing at 69.5 °C for 30 s, extension at 72 °C for 25 s, final extension at 72° C for 10 min, and cooling at 4 °C until the reaction stopped.

To confirm the efficiency and quality of the PCR reaction, aliquots of the amplification product were analyzed on a 1% agarose gel stained with ethidium bromide and the bands were visualized under ultraviolet (UV) light.

Purification of the DNA fragment of interest was carried out directly from the band visualized on the gel. A 35 µL aliquot of the PCR product was applied to 1% agarose gel for subsequent purification of the *Bhl* gene using a QIAquick Gel Extraction Kit. To confirm the absence of contamination in the sample, reagents, and equipment, a negative control was used for the reaction. The negative control contained phosphate deoxyribonucleotides (dNTPs: ATP, CTP, GTP, and TTP), Phusion Hot Start II High-Fidelity DNA Polymerase, magnesium ions (Mg^{2+}), forward and reverse primers, and the appropriate buffer for the enzyme (Buffer 5X HF). To confirm the size/position of the specific band of the *Bhl* gene, 5 µL of the same PCR product was also applied separately as a positive control. After quantification of the purified product on a NanoDrop spectrophotometer, a new run on a 1% agarose gel was conducted to confirm the size and quality of the purified DNA fragment.

Since the DNA polymerase produces "blunt ends", the *BhL* gene was adenylated at the 5′ ends to obtain cohesive ends and complementary to the 3′-T overhang of the pGEM-T Easy plasmid and, thus, allow the proper insertion of the *Bhl* gene into the vector. The adenylation reaction was performed using Taq DNA polymerase, in a reaction with a final volume of 10 µL, incubated at 72 °C for 60 min.

After adenylation, cloning was carried out in the pGEM T-Easy amplification vector. The binding reaction between the insert and the vector was carried out in a final volume of 10 µL and a molar ratio (insert/vector) equal to 3:1. Furthermore, a positive control, using the insert provided by the manufacturer, and a negative control were prepared. Reactions were incubated for 1 h at room temperature and overnight at 4 °C.

2.3.4. Transformation of Electrocompetent Cells

Electrocompetent *Escherichia coli* (DH5α strain) was transformed by electroporation with the pGEM T-Easy plasmid containing the *Bhl* gene. After electroporation, the cells were resuspended in 2x YT medium without the addition of antibiotics, and then an aliquot of approximately 100 µL was plated in 2x YT medium supplemented with ampicillin (0.1 g/mL). The plates were incubated at 37 °C for 16 h for cell growth in a bacteriological oven. In addition to the electroporation of electrocompetent cells with the recombinant plasmid, the pUC18 plasmid without an insert was used as a positive control and electroporated bacteria without the addition of any DNA sample as a negative control.

To identify colonies of bacteria transformed with the recombinant plasmid containing the *Bhl* gene, colony PCR was performed. At this stage, three colonies were randomly chosen to perform PCR and replated in 2x YT medium supplemented with ampicillin (0.1 g/mL) for isolation. The program used for colony PCR was the same as that described for gene amplification, according to Section 2.3.3.

2.3.5. Purification of Plasmid DNA

Colonies containing the *Bhl* gene were grown in a 2x YT liquid medium supplemented with ampicillin (0.1 g/mL) for the subsequent extraction of plasmid DNA. After 16 h of

incubation, the cells were pelleted by centrifugation, and their plasmids were purified by small-scale plasmid extraction using the Plasmid Miniprep Kit. After plasmid extraction, the purified DNA was quantified using a NanoDrop spectrophotometer. The presence and integrity of DNA were also analyzed through electrophoresis in a 1% agarose gel stained with ethidium bromide.

2.3.6. Sequencing

Aliquots of 10 µL of purified plasmid DNA in concentrations ranging from 150 to 300 ng/µL were sent to the René Rachou Institute—Fundação Osvaldo Cruz (Fiocruz, Brazil) for sequencing, which was conducted in hexaplicate. The method used was Sanger sequencing, and the external primers used were M13/pUC forward and M13/pUC reverse primers. The six repeats were aligned using the MultiAlin program (http://multalin.toulouse.inra.fr/multalin/, accessed on 1 February 2024), and all discrepancies in relation to the position of the nucleotides were visually checked and corrected individually using the Bioedit Sequence Alignment Editor, version 7.2.5 (https://bioedit.software.informer.com/7.2/, accessed on 29 January 2024) [34]. After obtaining the consensus sequence, it was deposited in GenBank and bioinformatic tools were used to predict the in silico, physicochemical, structural, and evolutionary properties of the BhL.

2.4. In Silico Characterization of the Lectin Associated with the Bhl Gene

2.4.1. Search for Homologous Sequences

The consensus genetic sequence obtained from sequencing was subjected to the BLAST algorithm to search for homologous nucleotide and protein sequences to the BhL, deposited and available in a database. The search for homologous nucleotide and protein sequences was initially carried out using the BLASTn program (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome, accessed on 1 February 2024), using the default parameters. Translation of the nucleotide sequence into a protein sequence was performed using the Translate tool (https://web.expasy.org/translate/, accessed on 1 February 2024). The largest open reading frame (ORF) was selected, starting with methionine and ending as a stop codon (Frame 2, 5′-3′ direction). The protein sequence obtained was used in FASTA format to search for homologous sequences using the BLASTp program (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins, accessed on 1 February 2024), with the standard search set indicated by the algorithm (Enter Query Sequence—FASTA format, Standard databases (nr, etc.), Algorithm—blastp/protein-protein BLAST).

2.4.2. Global Alignment between Homologous Sequences

To identify conserved regions and similarities and differences between sequences, protein sequences from *B. forficata* (APD76156.1), *B. variegata* (ABQ45362.1), *B. purpurea* (P16030.2), and *B. ungulata* (ABD19775.1), which showed greater similarity in BLASTp, were analyzed and compared to the BhL sequence using the Multialin program (http://multalin.toulouse.inra.fr/multalin/, accessed on 1 February 2024). The global alignment of homologous protein sequences also allowed for the determination of the presence of catalytic sites and conserved residues between BhL and the *Bauhinia* species selected in the analysis.

2.4.3. Analysis of the Physicochemical Properties and Characterization of Amino Acids

For the physicochemical characterization of the BhL protein, the computational tool ProtParam (https://web.expasy.org/protparam/, accessed on 2 February 2024) was used, which is part of the ExPASy server. The FASTA format of the protein sequence was also submitted to the ColorSeq platform (https://npsa.lyon.inserm.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_color.html, accessed on 2 February 2024) to characterize its amino acids.

2.4.4. Prediction of Post-Translational Modifications

Lectin protein sequences were also analyzed to predict post-translational modifications. The YinOYang 1.2 program (https://services.healthtech.dtu.dk/service.php?YinOYang-1.2, accessed on 2 February 2024) was used to predict possible O-glycosylation sites, and the NetPhos 3.1 web server (https://services.healthtech.dtu.dk/service.php?NetPhos-3.1, accessed on 2 February 2024) was used to predict phosphorylation sites.

2.4.5. Characterization of Functional Domains

The search for functional domains was carried out using InterProScan 100.0 software (https://www.ebi.ac.uk/interpro/search/sequence/, accessed on 3 February 2024).

2.4.6. Search for Signal Peptides

The search for the signal peptide was performed using SignalP 6.0 software (https://services.healthtech.dtu.dk/services/SignalP-6.0/, accessed on 7 June 2024) using the default parameters.

2.4.7. Prediction of Transmembrane Domains and Structural Characterization

The prediction of transmembrane helices in the protein was performed using the DeepTMHMM server (https://dtu.biolib.com/DeepTMHMM, accessed on 3 February 2024), version 1.0.24, using the program's default parameters.

The secondary and tertiary structures were predicted using the computational tools GOR4 (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html, accessed on 3 February 2024) and the AlphaFold3 server (https://alphafoldserver.com/, accessed on 7 June 2024), respectively.

3. Results

3.1. Obtaining the Nucleotide Sequence of B. holophylla Lectin

Electrophoretic analysis indicated the presence of intact genomic DNA at a concentration of 1565.8 ng/µL. To optimize *Bhl* gene amplification by PCR, a gradient with 12 distinct annealing temperature values was considered in a range that varied from 66.5 to 72 °C, with a half-degree scale between them. In this way, it was possible to observe the amplification of a fragment of approximately 900 base pairs, which corresponds to the expected size of the gene that encodes BhL. The best annealing temperature for the primers and the DNA polymerase activity was 69.5 °C. After purification, the *Bhl* gene showed a yield of 7.6 ng/µL.

Electrocompetent *E. coli* DH5α bacteria were transformed with the cloning product and plated on a 2x YT medium containing ampicillin. Colony growth indicated the incorporation of pGEM-T Easy plasmids. Amplified fragments of one colony had a size similar to that expected for the *Bhl* gene (Figure 1).

The cloning was confirmed by *Bhl* gene sequencing, which revealed an 867 bp nucleotide sequence. The *Bhl* gene sequence was submitted and deposited in the GenBank sequence repository (http://www.ncbi.nlm.nih.gov/genbank, accessed on 15 March 2024) under accession number PP501329. In silico translation of the nucleotide sequence revealed a protein consisting of 289 amino acids.

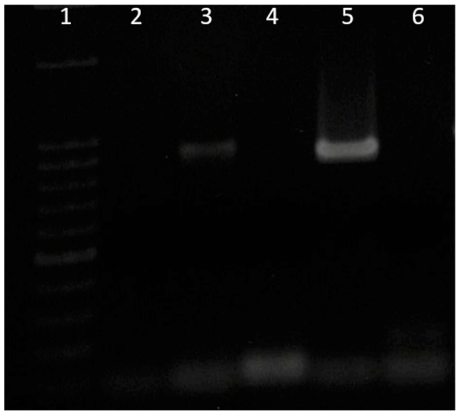


Figure 1. Product of the *Bhl* gene amplification. (1) 100 bp DNA Ladder; (2) negative control; (3) positive control (application of 5 µL of 5 ng/µL *Bhl* gene); (4–6) amplification product of the *Bhl* gene (approximately 900 bp) from *E. coli* DH5α colonies transformed with the cloning product (pGEM-T easy + *Bhl* gene).

3.2. In Silico Characterization of the Lectin Associated with the Bhl Gene

3.2.1. Search for Homologous Sequences and Conserved Regions

The alignment performed by BLASTn indicated a high similarity between the *Bhl* gene and nucleotide sequences that encode lectins expressed by *Bauhinia* species deposited in the database (Table 2). This tool revealed identity percentages of 95.06, 92.01, 89.84, and 87.56% for lectins from *B. forficata*, *B. ungulata*, *B. variegata*, and *B. purpurea*, respectively.

Table 2. Search for identity of the nucleotide sequence of *Bhl* gene with other proteins in NCBI database. Results obtained by BLASTn program.

Accession (NCBI)	Description	Max Score	Total Score	Query Cover	E Value	Per. Identity
KX907616.1	*Bauhinia forficata* lectin	1365	1365	100%	0.0	95.06%
DQ372702.1	*Bauhinia ungulata* mRNA, partial cds	1079	1079	88%	0.0	92.01%
EU596376.1	*Bauhinia variegata* lectin I (bvl) gene, bvl-2 allele, complete cds	1114	1114	100%	0.0	89.84%
D12481.1	*Bauhinia purpurea* mRNA, complete cds	1000	1000	100%	0.0	87.56%

In addition, the BLASTp 2.15.0 software also showed a high identity between the protein sequence of BhL and lectins from *B. forficata* (90%), *B. ungulata* (85.27%), *B. variegata* (79.04%), and *B. purpurea* (78.01%). In addition to *Bauhinia* species, *Griffonia simplicifolia* showed 69.83% identity and 83% alignment with the *Bhl* gene (Table 3).

Table 3. Search for identity of the amino acid sequence of BhL with other proteins in NCBI database. Results obtained by BLASTp program.

Accession (NCBI)	Description	Max Score	Total Score	Query Cover	E Value	Per. Identity
APD76156.1	*Bauhinia forficata* lectin	521	521	100%	0.0	90.00%
ABD19775.1	*Bauhinia ungulata* lectin	436	436	88%	4e−152	85.27%
ABQ45362.1	*Bauhinia variegata* lectin I precursor	452	452	100%	9e−158	79.04%
P16030.2	RecName: Full = Lectin; Flags: Precursor [*Bauhinia purpurea*]	440	440	100%	3e−153	78.01%
P24146.3	RecName: Full = Lectin-4; AltName: Full = GS4; AltName: Full = Lectin IV [*Griffonia simplicifolia*]	338	338	83%	1e−113	69.83%

3.2.2. Global Alignment between Homologous Sequences

Lectin sequences from *B. forficata*, *B. variegata*, *B. purpurea*, and *B. ungulata* were aligned simultaneously with the *Bhl* gene on the MultiAlin platform (Figure 2).

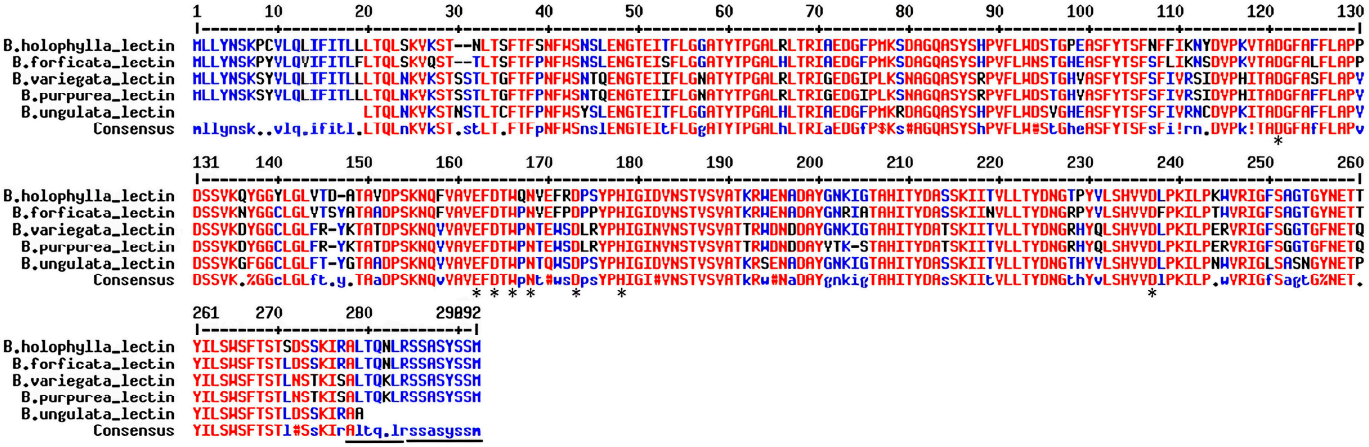


Figure 2. Global alignment of BhL sequence with proteins that showed high homology. Results generated by the MultiAlin platform, where red represents strong homology (above 90%), blue represents weak homology (below 50%), and black represents very low homology or no consensus. First line: *B. holophylla*_lectin; Second line: *B. forficata*_lectin; Third line: *B. variegata*_lectin; Fourth line: *B. purpurea* _lectin; Fifth line: *B. ungulata*_lectin; Sixth line: consensus sequence. Important residues for the active conformation of the carbohydrate recognition domain (CRD) are marked manually with an asterisk (*) and conserved cleavage sites with a dash (_). In the consensus sequence, (!) means that there was an exchange of Isoleucine (I) for Valine (V); ($) means that Leucine (L) was replaced by Methionine (M); (%) means that there was an exchange of Phenylalanine (F) for Tyrosine (Y); (#) means the exchange for one of the amino acids: Asparagine (N), Aspartate (D), Glutamine (Q), Glutamate (E), Asparagine (B), or Glutamine (Z).

The results indicated the presence of many regions with high identity between the sequences, with a more significant absence of consensus at the N- and C-terminal ends. This was mainly due to the shorter length of the *B. ungulata* sequence, which consequently led to a smaller number of amino acids in the final protein. Two cleavage sites for the Arg-C proteinase were predicted in BhL, namely ALTQNLR (residues 278–284) and SSASYSSM (residues 285–292) in the C-terminal region of the protein sequence. The aspartate residue (Asp237), which is considered important for the formation of the metal binding site (MBS) in legume lectins, remained conserved in all aligned *Bauhinia* species. Multiple sequence alignment also showed that BhL had highly conserved catalytic residues, such as Asp121, Gly142, Trp166, and Asn168, which constitute the CRD, and the residues Glu162, Asp164, Trp166, Asn168, Asp173, and His 178, which make up the site bonding to metals. The presence of these catalytic residues brings BhL functionally close to the lectins present in species of the Caesalpinioideae subfamily.

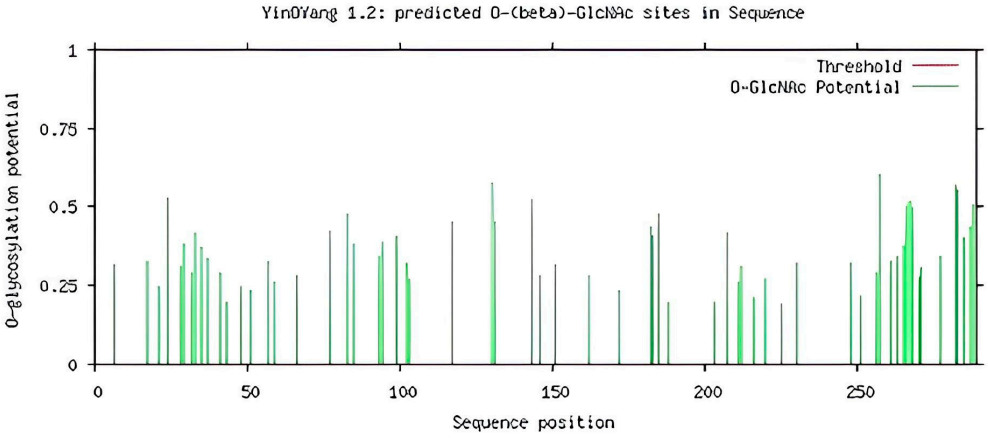
3.2.3. Analysis of Physicochemical Properties and Characterization of Amino Acids

Physicochemical analysis confirmed the presence of 289 amino acids in BhL, estimating a molecular weight of 31.9 kDa and a theoretical isoelectric point of 5.79. The ProtParam program also provided a prediction of the estimated half-life of the lectin in different cell types. In vitro, the protein had a half-life of 30 h in mammalian reticulocytes, and in vivo, it was 20 h in yeast and 10 h in *Escherichia coli*.

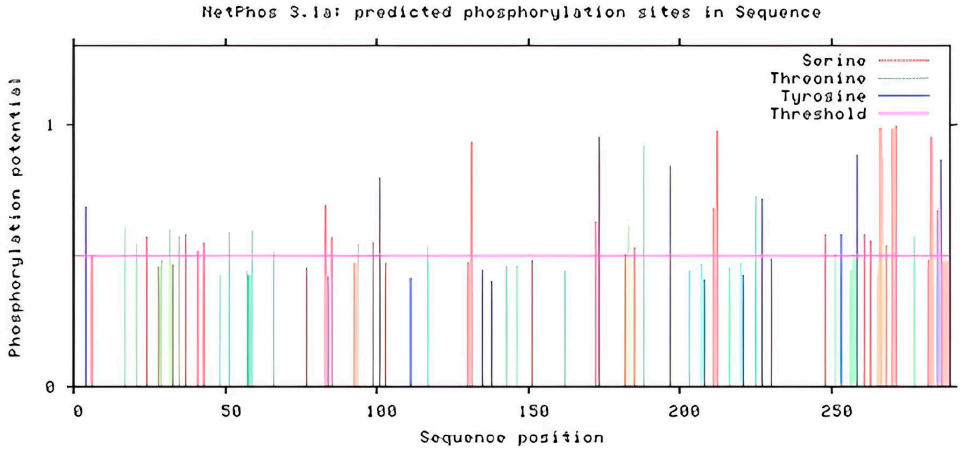
ColorSeq software revealed the presence of 40 aromatic residues and 78 hydroxyl residues, in addition to a very equal relationship between positive (25) and negative (24) residues throughout the sequence of 289 amino acids present in BhL, with 88 amino acids possessing hydrophobic residues and 130 amino acids with hydrophilic residues.

3.2.4. Prediction of Post-Translational Modifications

The analyses revealed that BhL possibly undergoes phosphorylations and glycosylations at specific sites throughout its protein sequence (Figure 3). Analyses carried out on the YinOYang platform predicted that serine and threonine amino acids undergo glycosylation throughout the BhL protein (Figure 3a), and the NetPhos program indicated that the residues most likely to be phosphorylated were serine, threonine, and tyrosine (Figure 3b).



(a)



(b)

Figure 3. Prediction of post-translational modifications. (**a**) O-β-GlcNAc attachment sites generated by YinOYang platform; (**b**) serine, threonine, or tyrosine phosphorylation sites obtained by NetPhos tool. For (**a**), the threshold is equal to zero.

3.2.5. Characterization of Functional Domains and the Presence of Signal Peptides

Based on predictive models, InterProScan 100.0 software recognized conserved protein domains in BhL, which brought it closer to lectin families from other legume species, especially Concanavalin A (ConA) (Family and Homologous Superfamily). The tool also indicated high similarity between the Leguminosae β-chain (Domains and Representative Domains) and the region located between amino acids 156 and 162 of BhL (Binding Site). Finally, the prediction signaled the presence of a sequence of 28 amino acids at the N-terminal of the lectin, with a high hydropathicity index and conceptualized as a signal peptide (Other Features), as well as the presence of important residues for the constitution of interaction and binding sites characteristic of lectins (Residues) (Figure 4).

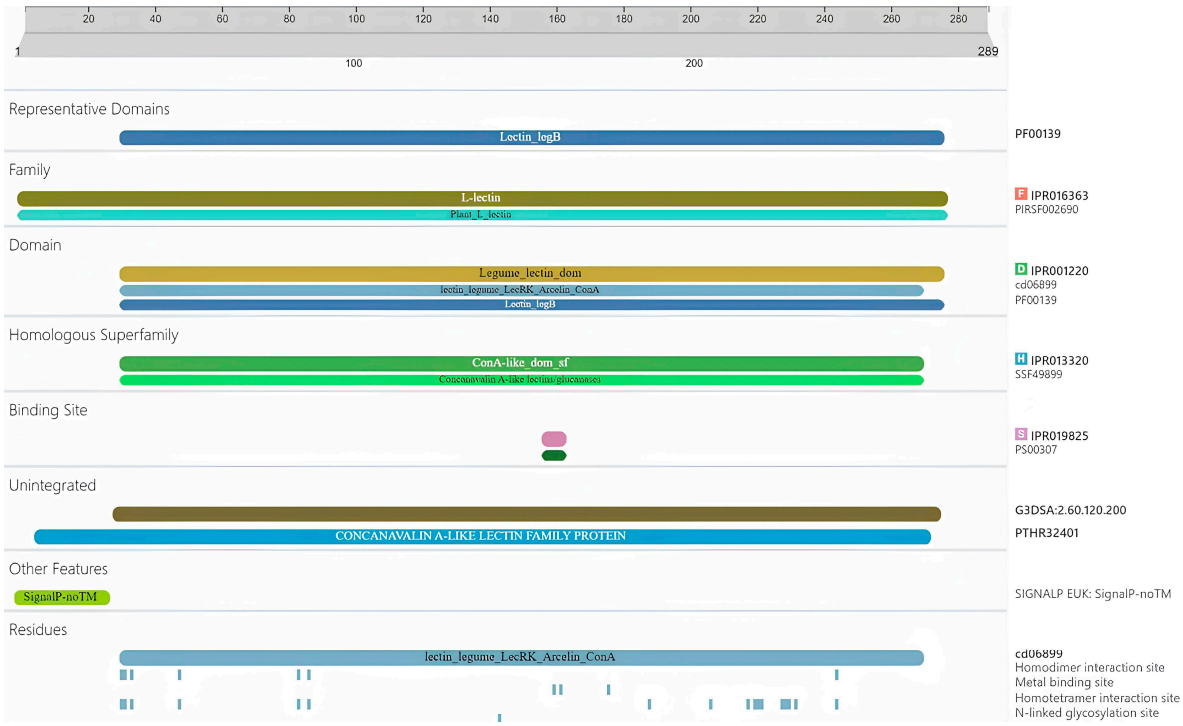


Figure 4. Search for functional domains, important sites, and families for BhL. Results obtained by InterProScan tool.

The presence of signal peptides was confirmed using the SignalP platform, which indicated a 96.44% probability of signal peptides in the N-terminal portion of BhL. The cleavage site was identified between amino acids 28 and 29 of the protein sequence (Figure 5).

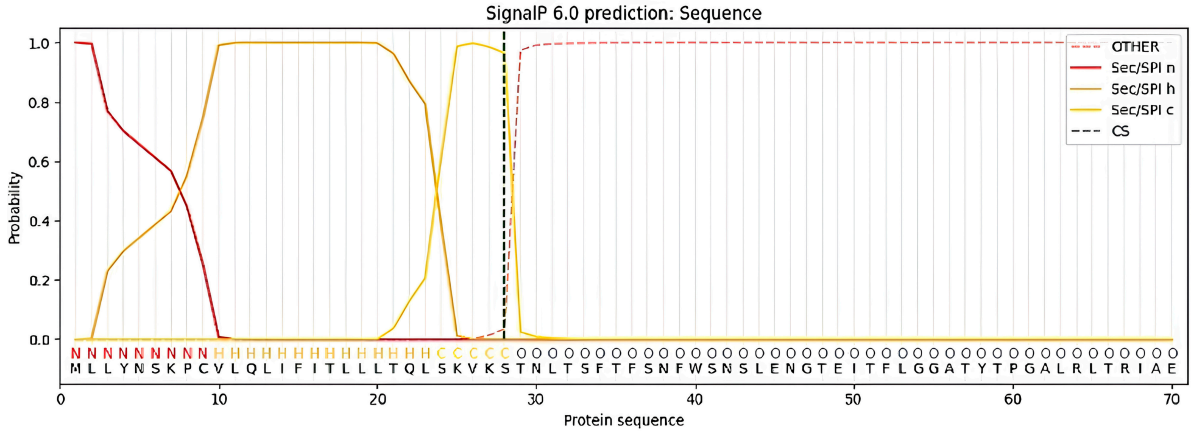


Figure 5. Analysis in SignalP 6.0 program. Prediction of signal peptide presence in BhL, where (—) Sec/SPIn: "standard" secretory signal peptides transported by the Sec translocon and cleaved by Signal Peptidase I (Lep)—n-terminal region; (—) Sec/SPIh: "standard" secretory signal peptides transported by the Sec translocon and cleaved by Signal Peptidase I (Lep)—center hydrophobic region; (—) Sec/SPIc: "standard" secretory signal peptides transported by the Sec translocon and cleaved by Signal Peptidase I (Lep)—c-terminal region; (---) probable cleavage site for the signal peptide; (---) Other.

3.2.6. Prediction of Transmembrane Domains and Structural Characterization

Figure 6 shows the probability of finding intracellular, transmembrane, or extracellular portions in BhL. Analyses carried out with DeepTMHMM 1.024 software pointed to the absence of transmembrane helices in BhL, as well as a predominantly extracellular membrane

topology. The noise observed (in orange color) at the beginning of the plot suggests the presence of a signal peptide in the N-terminal portion of the protein. These results, added to the signal peptide prediction, allowed us to infer that BhL may be a protein that will be directed to the classical intracellular transport pathway, after translocation to the lumen of the endoplasmic reticulum.

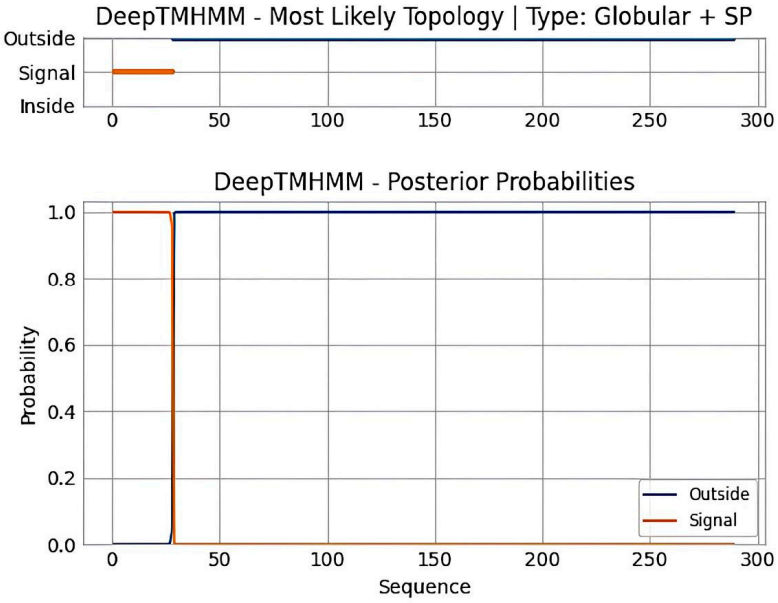


Figure 6. Search for the presence of transmembrane helices in BhL by DeepTMHMM toll. (—) probability to be an extracellular portion; (—) probability to be a signal peptide.

Secondary structure prediction was performed using GOR4 software. The analysis indicated percentages of 10.38% for α-helices, 29.41% for β-sheets, and 60.21% for disordered regions. Thus, of the 289 amino acids present in the protein, 30, 85, and 174 residues were related to α-helix, β-sheet, and disordered regions, respectively (Figure 7).

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MLLYNSKPCVLQLIFITLLLTQLSKVKSTNLTSFTFSNFWSNSLENGTEITFLGGATYTPGALRLTRIAE
ccccccccceeeeehhhhhhhhhcccccccceeeeeecccccccccccceeeeecccccccchhhhhhhc
DGFPMKSDAGQASYSHPVFLWDSTGPEASFYTSFNFFIKNYDVPKVTADGFAFFLAPPDSSVKQYGGYLG
ccccccccccccccccceeeecccccccccccccceeecccccccccccceeeecccccccccccceee
LVTDATAVDPSKNQFVAVEFDTWQNVEFRDPSYPHIGIDVNSTVSVATKRWENADAYGNKIGTAHITYDA
eeecccccccccccceeeeeecccccccccccccccceeeeeccceeeeeehhhhhhcccccccceeeeeccc
SSKIITVLLTYDNGTPYVLSHVVDLPKILPKWVRIGFSAGTGYNETTYILSWSFTSTSDSSKIRALTQNL
ccceeeeeeecccccceeeecccccccccccceeeeecccccccceeeeeeeeeecccccchhhhhhhhcc
RSSASYSSM
cccceeeec

Sequence length :   289
```

Figure 7. Prediction of the secondary structures by GOR tool, where (h) represents α-helical regions; (e) represents regions of β-sheets; (c) represents regions of random coils.

The AlphaFold3 server used advanced modeling techniques to predict the three-dimensional structure of BhL with high accuracy. Figure 8 shows the 3D structure of the BhL, with a predicted template modeling (pTM) score equal to 0.87.

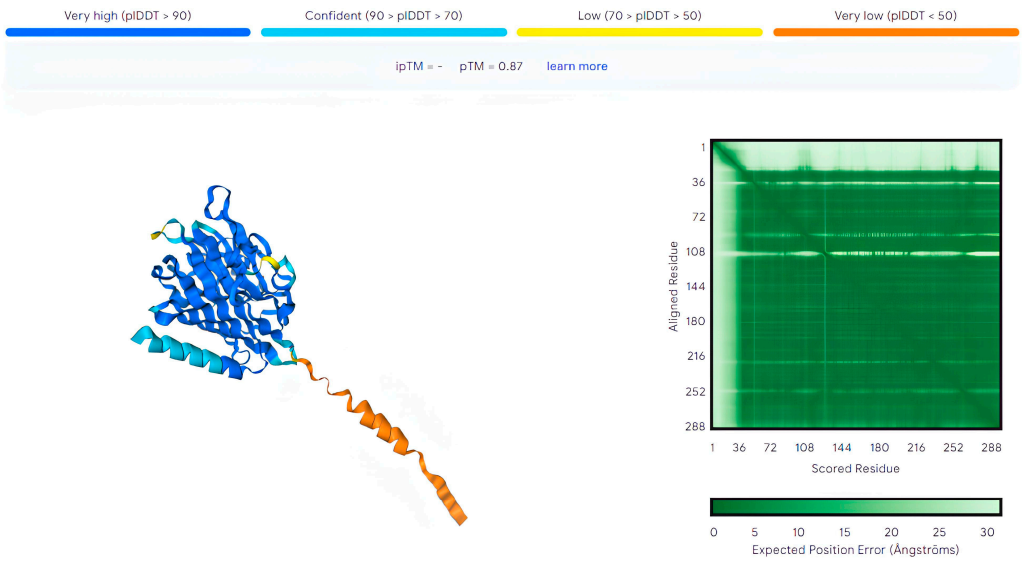


Figure 8. Schematic representation of 3D structure of *B. holophylla* lectin by AlphaFold3 server. Local Distance Difference Test (lDDT) is a superposition-free score that evaluates local distance differences of all atoms in a model, with the objective of estimating the confidence of the predicted model. Dark blue: regions with pIDDT greater than 90 (Very high); light blue: regions with pIDDT between 70 and 90 (Confident); yellow regions with pIDDT between 50 and 70 (Low); orange: regions with pIDDT lower than 50 (Very low). The predicted template modeling (pTM) score and the interface predicted template modeling (ipTM) score evaluate the prediction of the structure. In green, the predicted aligned error (PAE) plot is the estimate of the error in the relative position and orientation between two tokens in the predicted structure.

4. Discussion

Technological advances in genetics and molecular biology associated with the use of bioinformatics tools have allowed the sequencing of a large number of genes, as well as the prediction of physicochemical characteristics and biological functions of several proteins [20]. However, although legume lectins have been investigated for many decades, information on recombinant lectins from *Bauhinia* species is still scarce, with few proteins of the genus isolated, characterized, and with nucleotide sequences duly deposited in public databases [6]. Among the more than 300 species cataloged and taxonomically classified within the *Bauhinia* genus, less than 5% of them had their lectins isolated, such as lectins from *B. purpurea*, *B. monandra*, *B. pentandra*, *B. variegata*, *B. bauhinioides*, *B. forficata*, and *B. ungulata* [35–41]. For *B. holophylla* lectin, no studies involving the characterization at the structural and functional level or of its possible isoforms have been found to date.

In the present study, several bioinformatics tools were used for the physicochemical, structural, and functional characterization of a lectin isoform from *B. holophylla*, sequenced from genetic material extracted from plant callus. The molecular biology techniques used allowed the identification of a nucleotide sequence of 867 bp, whose in silico translation resulted in a protein composed of 289 amino acids. Since there were no database records of lectin coding sequences expressed in *B. holophylla*, the sequence was submitted and deposited in the GenBank sequence repository (http://www.ncbi.nlm.nih.gov/genbank, accessed on 15 March 2024) under accession number PP501329.

Initially, analyses carried out on the BLAST platform revealed a high identity (greater than 85%) between BhL and lectins from *B. forficata*, *B. variegata*, *B. purpurea*, and *B. ungulata* (Table 3), which corroborated the hypothesis that the gene family of legume lectins is represented by a large family of homologous proteins that arose from a single common ancestral gene [42]. Furthermore, the results suggested a considerable percentage of identity (69.83%) between BhL and the lectin from *Griffonia simplicifolia*, which is also a species of medicinal legume belonging to the Caesalpinaceae subfamily and originating from

tropical regions of Africa [43]. Studies involving the analysis of nucleotide or protein sequences of plant lectins have indicated the presence of evolutionarily conserved regions that may contain domains associated with the carbohydrate binding and recognition site, which justifies the similarity found between species of different legume genera [44]. Since the genetic code is degenerate, the percentages of identities predicted between proteins (BLASTp) would be greater than those identified between gene sequences (BLASTn). However, the discrepancy observed between the results presented in Tables 2 and 3 suggests that the accessions referring to gene and protein sequences, used by the BLASTn and BLASTp programs, respectively, in the comparative analyses, are not coincident.

The presence of many regions highlighted in red in the global alignment of protein sequences (Figure 2) confirmed the high level of consensus between BhL and the other *Bauhinia* lectins recovered from BLASTp. A more significant discrepancy was noted at the N- and C-terminal ends containing regions with less identity. However, sequence ABD19775.1, described in GenBank as a lectin from *B. ungulata*, may have a shorter amino acid sequence or have been deposited incompletely. It is known that public databases are still under construction, and errors in deposited sequences are to be expected given the complexity of the entire process involved, starting with genome sequencing, data assembly, annotations, and sequence curation [45]. Sequence alignment also indicated that BhL had two cleavage sites for Arg-C proteinase, namely ALTQNLR (residues 278–284) and SSASYSSM (residues 285–292). These sites allow the cleavage of peptides located in the C-terminal position in relation to the arginine residue. Pinto et al. [15], in a study carried out with *B. forficata*, also indicated the presence of these cleavage sites and their importance for the processing of the mature protein. The aspartate residue (Asp237), which was shown to be conserved in all *Bauhinia* species, is also present in the mature ConA protein (Asp208). In ConA, this amino acid acts as an important bridge for the binding of calcium ions and transition metal ions at the MBS present in legume lectins [46]. In addition to this aspartate residue, in ConA, studies suggest the need for an asparagine residue and an aromatic amino acid composing what is known as a "conserved triad", for the perfect functioning of the CRD. In lectins from the Caesalpinioideae subfamily, the CRD showed high similarity, even in lectins with different carbohydrate affinities. The CRD of lectins from the Caesalpinioideae subfamily consists of residues Asp121, Gly142, Trp166, and Asn168, which were identified in BhL. The MBS is made up of the residues Glu162, Asp164, Trp166, Asn168, Asp173, and His 178, which are also present in BhL [3,47].

In general, the classification of plant lectin-encoding genes into families is based on the similarity between protein sequences, the observation of evolutionary relationships, and conserved carbohydrate recognition domains [48]. The Clustal Omega annotation tool used to establish evolutionary relationships between BhL and other lectins from *Bauhinia* species indicated a more recent common ancestry between *B. forficata* lectin genes and the *BhL* gene. The alignment of the two sequences demonstrated that there were 28 amino acid substitutions, of which 11 were conservative, 7 were semi-conservative, and 10 were radical. Furthermore, there was a deletion of an amino acid in BhL at position 145 of the protein sequence compared to the *B. forficata* lectin. The preservation of intact catalytic sites and conserved amino acids in the BhL sequence, as well as in other *Bauhinia* lectins, underscores the potential significance of these structural elements in facilitating the protein's activities and functions. This observation implies that the architecture of these sites and residues likely plays a crucial role in maintaining the protein's functional integrity [48,49].

Variations in the evolutionary rates of protein families involve individual, functional, and structural factors that can facilitate or hinder the occurrence of permanent substitutions in the nucleotide sequence [49,50]. Among these factors, the level of gene expression is considered decisive for the accumulation of evolutionary modifications in proteins. Therefore, families of proteins that are broad and highly expressed evolve more slowly and, consequently, may share conserved regions throughout their sequences [51]. Plant lectins represent a group of proteins with a cosmopolitan distribution, with an emphasis on lectins from the legume family, which emerge as the best-described and -investigated

group [6]. However, genes encoding plant lectins are normally constitutive and expressed continuously under normal conditions of plant development [1,4].

Physicochemical analyses indicated that BhL had 289 amino acids, a molecular weight of 31.9 kDa, and a pI of 5.79. Molecular weights ranging between 26 and 33 kDa have also been observed for lectins expressed in other *Bauhinia* species [6]. The theoretical pI of the BhL indicates a negative charge on the protein and an acidic nature. For comparison, physicochemical analyses were performed for the four *Bauhinia* lectin protein sequences analyzed by MSA in the ProtParam program [52]. The predictions showed that *Bauhinia* lectins have very similar physicochemical characteristics (Table S1). A greater discrepancy was noted in relation to the predicted pI, which does not compromise the soluble character of the proteins, as indicated by the hydropathi-city index (GRAVY). However, changes in pI can influence other parameters such as, for example, subcellular localization or interactions established by proteins [53]. Complementarily, ColorSeq estimated the presence of 40 aromatic residues, 78 hydroxyl residues, 88 hydrophobic residues, and 130 hydrophilic residues in BhL, pointing to a low hydropathicity index and hydrophilic character of the protein. Lectins share many molecular characteristics, such as a high number of hydroxyl residues [47]. Pinto et al. [15] showed that in *Bauhinia forficata,* the amino acid composition of the BfL-II lectin presented percentages of 34.35% (90 residues) and 35.50% (93 residues) of polar and hydrophobic amino acids, respectively. The computational tools YinOYang and NetPhos indicated that BhL possibly underwent post-translational processing events. Plant lectins are generally synthesized as a precursor protein that, after cleavage of a signal peptide located in the N-terminal position, undergoes post-translational modifications to become a mature lectin [9,54]. Studies have revealed that this processing may involve proteolytic cleavage of the precursor chain, glycosylations, cuts in the C-terminal region, the removal of the covalent carbohydrate, and ligation between the C- and N-terminal ends [15,42,44,46]. Cagliari [6] reported that lectins from *B. forficata*, *B. monandra*, *B. purpurea*, *B. ungulata,* and *B. variegata* are glycoproteins, which undergo pro-glycosylation processes. In this way, such analyses make important predictions about mature proteins, which is extremely useful for designing future experiments and in vitro tests.

The prediction of conserved domains and potential functions revealed that BhL had domains representative of the legume lectin family, which reinforces the results found in BLASTp. The program also indicated the relationships of BhL with domains of the ConA-like protein superfamily (InterPro: IPR013320), which represents a widely studied plant hemagglutinin found in *Canavalia ensiformis* [6,55]. According to Van Holle et al. [48], the protein domains of lectins are evolutionarily well preserved among structurally conserved taxa and, often, functionally independent of the rest of the protein. These domains evolved from shorter structural units, such as repeats, or from the association of small folding motifs. The presence of a metal ion binding site (Ca2+ and Mn2+) and a signal peptide at the N-terminal end of the BhL were also estimated. Many legume lectins are classified as metalloproteins and have binding sites in their structure for calcium and manganese ions (or other transition metals), essential for the performance of their biological activities [3]. Although metal ions do not directly bind to carbohydrates, their presence is crucial for lectin to achieve its quaternary structure and functional conformation [55]. In ConA, manganese first binds to the S1 site of the protein, which leads to the formation of a second site for calcium ion binding (S2 site). Only when both S1 and S2 sites are occupied does the ConA lectin assume a conformation that allows it to recognize carbohydrates [54,56]. Thus, the results allow us to infer not only the shared evolutionary origins but also that the molecular function of BhL is possibly associated with carbohydrate recognition activity. The presence of signal peptides was also predicted by the SignalP 6.0 software, indicating a cleavage site between amino acids 28 and 29 of the BhL sequence. Signal peptides correspond to short amino acid sequences (16–30 residues), generally located in the N-terminal portion of the protein. For secretory and membrane proteins, these signal sequences mediate the targeting of the nascent polypeptide chain to the endoplasmic reticulum, where protein elongation proceeds [57]. Legume lectins are commonly synthesized with a signal sequence

at the N-terminal end, which is responsible for guiding them to the secretory pathway [48]. In *B. forficata*, in silico studies pointed to the presence of a signal peptide formed by the first 28 amino acids of the protein sequence, resulting in a mature protein of 262 amino acids [15]. The absence of transmembrane loops and the soluble nature of BhL, combined with the presence of a signal peptide, indicated that BhL is secreted by the plant. A study of the biosynthesis of several legume lectins demonstrates that after their synthesis in the endoplasmic reticulum, they are directed to the secretory pathway [1].

A crystallographic study demonstrated that the secondary and three-dimensional structures of legume lectins are very similar to each other and are characterized by a low level of α-helix arrangements compared to β-sheets [15]. According to Silva [40], the secondary structure of *B. ungulata* lectin presents 4% α-helix and 27% β-sheets compared to 10% α-helices and 29% β-sheets observed for BhL. In *B. forficata*, spectrophotometric analyses indicate the presence of 19% α-helices and 27% β-sheets [39]. A protein structure characterized by fewer α-helices and more β-sheets may exhibit increased stability due to the extensive hydrogen bonding in β-sheets, but it could also be more prone to aggregation and amyloid formation, potentially affecting its folding dynamics and making it more vulnerable to denaturation under certain conditions [58–60]. The prevalence of β-sheets in the secondary structure of legume lectins, such as those found in *Bauhinia holophylla*, not only contributes to their stability and resistance to proteolysis, as highlighted by Perczel et al. [59] but also plays a pivotal role in defining their functional architecture. This structural motif, known as the "jelly roll fold" observed in legume lectins like Concanavalin A and inferred in *Bauhinia forficata* lectins (BfL-I and BfL-II) by Pinto et al. [15], facilitates carbohydrate binding by providing a suitable conformation for the sugar-binding pocket. This conserved β-sandwich structure, characterized by antiparallel β-sheets and hydrophobic cores, underscores the significance of β-sheet predominance in dictating legume lectins' function and substrate interactions within the cell [61]. Regarding the three-dimensional structure of BhL, AlphaFold predicted a 3D structure with reliable metrics, proposing a useful model for further in-depth studies on the biological function and biotechnological potential of BhL [62].

In today's rapidly advancing era of sequencing techniques and the explosion of genomic data, developing computational solid methods for predicting protein structures is crucial. This is important for effective functional annotation because obtaining experimental structural data, especially for lesser-studied proteins, is challenging. Therefore, relying on initial in silico analyses is becoming increasingly crucial [63,64]. Consequently, alongside the evolution of prediction methodologies, there is a parallel need for tools to validate the accuracy of these computational models. This validation process is instrumental in refining protein structures, as recent research underscores it as a significant bottleneck affecting both the quality and confidence levels of predicted protein structures [64].

Looking ahead, the predictions outlined in our manuscript will be further validated through molecular modeling of the BhL lectin. This forthcoming analysis aims to elucidate the intricate interactions between the amino acids within the catalytic sites and the specific substrate to which the CRD binds.

5. Conclusions

In this study, a lectin isoform from *B. holophylla* was sequenced and characterized using bioinformatics tools. The in silico analyses allowed for obtaining important information regarding its physicochemical characteristics, secondary and tertiary structures, and biosynthesis. Furthermore, computational biology tools allowed us to infer that the biological function of this isoform is related to the recognition of carbohydrates. The in silico characterization of new legume lectins can provide valuable data about this large family of homologous proteins and represents an important step towards planning future steps for the synthesis of recombinant proteins and assessment of their biological activities.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/horticulturae10070682/s1, Table S1: Physicochemical characterization of lectins expressed in Bauhinia species using the ProtParam program. pI—Isoelectric point; Gravy—Grand average of hydropathicity.

Author Contributions: Conceptualization, D.d.O.L. and A.H.F.C. Data curation, L.M.G.G.d.C. and D.d.O.L. Formal analysis, L.M.G.G.d.C. Funding acquisition, A.H.F.C. Investigation, L.M.G.G.d.C., D.d.O.L. and A.H.F.C. Methodology, L.M.G.G.d.C., D.d.O.L. and A.H.F.C. Project administration, A.H.F.C. Resources, D.d.O.L. and A.H.F.C. Supervision, A.H.F.C.; Validation, L.M.G.G.d.C. and D.d.O.L. Visualization, L.M.G.G.d.C., D.d.O.L. and A.H.F.C. Writing—original draft, L.M.G.G.d.C., D.d.O.L. and A.H.F.C. Writing—review and editing, L.M.G.G.d.C., D.d.O.L. and A.H.F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Processo 306581/2020-5) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) (APQ-00861-22 and APQ-04400-23). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES), Finance Code 001.

Data Availability Statement: The original contributions presented in the study are included in the article and Supplementary Material, further inquiries can be directed to the corresponding authors.

Acknowledgments: We thank Andreia Fonseca Silva for the botanical identification of this plant. The authors are grateful to the Universidade Federal de São João del-Rei (UFSJ), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for financial support. L.M.G.G.C is grateful to CAPES for granting the doctoral scholarship.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Van Damme, E.J.M. 35 years in plant lectin research: A journey from basic science to applications in agriculture and medicine. *Glycoconj. J.* **2022**, *39*, 83–97. [CrossRef] [PubMed]
2. Bellande, K.; Bono, J.J.; Savelli, B.; Jamet, E.; Canut, H. Plant lectins and lectin receptor-like kinases: How do they sense the outside? *Int. J. Mol. Sci.* **2017**, *18*, 1164. [CrossRef] [PubMed]
3. Cavada, B.S.; Pinto-Junior, V.R.; Osterne, V.J.S.; Oliveira, M.V.; Lossio, C.F.; Silva, M.T.L.; Bari, A.U.; Lima, L.D.; Souza-Filho, C.H.D.; Nascimento, K.S. Comprehensive review on Caelsalpinioideae lectins: From purification to biological activities. *Int. J. Biol. Macromol.* **2020**, *162*, 333–348. [CrossRef] [PubMed]
4. Tsaneva, M.; Van Damme, E.J.M. 130 years of plant lectin research. *Glycoconj. J.* **2020**, *37*, 533–551. [CrossRef] [PubMed]
5. Lagarda-Diaz, I.; Guzman-Partida, A.M.; Vazquez-Moreno, L. Legume lectins: Proteins with diverse applications. *Int. J. Mol. Sci.* **2017**, *18*, 1242. [CrossRef] [PubMed]
6. Cagliari, R.; Kremer, F.S.; Pinto, L.S. *Bauhinia* lectins: Biochemical properties and biotechnological applications. *Int. J. Biol. Macromol.* **2018**, *119*, 811–820. [CrossRef]
7. Vaz, A.M.S.F. *Bauhinia* in Lista de Espécies da Flora do Brasil. Jardim Botânico do Rio de Janeiro. Available online: https://floradobrasil.jbrj.gov.br/FB111840 (accessed on 22 November 2023).
8. Filho, V.C. Chemical composition and biological potential of plants from the genus Bauhinia. *Phytother. Res.* **2009**, *23*, 1347–1354. [CrossRef] [PubMed]
9. Lannoo, N.; Van Damme, E.J.M. Lectin domains at the frontiers of plant defense. *Front. Plant Sci.* **2014**, *5*, 397. [CrossRef]
10. Naithani, S.; Komath, S.S.; Nonomura, A.; Govindjee, G. Plant lectins and their many roles: Carbohydrate-binding and beyond. *J. Plant Physiol.* **2021**, *266*, 153531. [CrossRef]
11. Van Holle, S.; Van Damme, E.J.M. Signaling through plant lectins: Modulation of plant immunity and beyond. *Biochem. Soc. Trans.* **2018**, *46*, 217–233. [CrossRef]
12. Girão, D.K.F.B.; Cavada, B.S.; Pires, A.F.; Martins, T.V.; Franco, A.X.; Morais, C.M.; Nascimento, K.S.; Delatorre, P.; Silva, H.C.; Nagano, C.S.; et al. The galactose-binding lectin isolated from *Bauhinia bauhinioides* Mart. seeds inhibits neutrophil rolling and adhesion via primary cytokines. *J. Mol. Recognit.* **2015**, *28*, 285–292. [CrossRef]
13. Campos, J.K.L.; Araújo, C.S.F.; Araújo, T.F.S.; Santos, A.F.S.; Teixeira, J.A.; Lima, V.L.M.; Coelho, L.C.B.B. Anti-inflammatory and antinociceptive activities of *Bauhinia monandra* leaf lectin. *Biochim. Open* **2016**, *2*, 62–68. [CrossRef]
14. Lubkowski, J.; Durbin, S.V.; Silva, M.C.C.; Farnsworth, D.; Gildersleeve, J.C.; Oliva, M.L.V.; Wlodawer, A. Structural analysis and unique molecular recognition properties of a *Bauhinia forficata* lectin that inhibits cancer cell growth. *FEBS J.* **2017**, *284*, 429–450. [CrossRef] [PubMed]

15. Pinto, L.S.; Cardoso, G.; Kremer, F.S.; Woloski, R.D.S.; Dellagostin, O.A.; Campos, V.F. Heterologous expression and characterization of a new galactose-binding lectin from *Bauhinia forficata* with antiproliferative activity. *Int. J. Biol. Macromol.* **2019**, *128*, 877–884. [CrossRef]
16. Konozy, E.; Osman, M.; Dirar, A. Plant lectins as potent anti-coronaviruses, anti-inflammatory, antinociceptive and antiulcer agents. *Saudi J. Biol. Sci.* **2022**, *29*, 103301. [CrossRef]
17. Zhong, C.; Zhou, Z.; Zhang, Y.M.; Jia, S.R.; Sun, Z.; Dale, B.E. Integrating kinetics with thermodynamics to study the alkaline extraction of protein from *Caragana korshinskii* Kom. *Biotechnol. Bioeng.* **2014**, *111*, 1801–1808. [CrossRef] [PubMed]
18. Saw, H.S.; Sankaran, R.; Khoo, K.S.; Chew, K.W.; Phong, W.N.; Tang, M.S.Y.; Lim, S.S.; Zaid, H.F.M.; Naushad, M.; Show, P.L. Application of a liquid biphasic flotation (LBF) system for protein extraction from *Persiscaria Tenulla* leaf. *Processes* **2020**, *8*, 247. [CrossRef]
19. Raemaekers, R.J.M.; Muro, L.; Gatehouse, J.A.; Fordham-Skelton, A.P. Functional phytohemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris*. *Eur. J. Biochem.* **1999**, *265*, 394–403. [CrossRef] [PubMed]
20. Silva, F.F.; Gonçalves, D.B.; Lopes, D.O. The use of bioinformatics tools to characterize a hypothetical protein from *Penicillium rubens*. *Genet. Mol. Res.* **2020**, *19*, 18574. [CrossRef]
21. Mallick, P.; Kuster, B. Proteomics: A pragmatic perspective. *Nat. Biotechnol.* **2010**, *28*, 695–709. [CrossRef]
22. Sivashankari, S.; Shanmughavel, P. Functional annotation of hypothetical proteins—A review. *Bioinformation* **2006**, *1*, 335–338. [CrossRef]
23. Srinivasa, K.G.; Siddesh, G.M.; Manisekhar, S.R. Statistical modelling and machine learning principles for bioinformatics techniques, tools, and applications. In *Algorithms for Intelligent Systems*; Srinivasa, K.G., Siddesh., G.M., Manisekhar, S.R., Eds.; Springer: Singapore, 2020; pp. 3–9.
24. Ijaq, J.; Malik, G.; Kumar, A.; Das, P.S. A model to predict the function of hypothetical proteins through a nine-point classification scoring schema. *BMC Bioinform.* **2019**, *20*, 14. [CrossRef] [PubMed]
25. Helmick, H.; Jain, A.; Terashi, G.; Liceaga, A.; Bhunia, A.K.; Kihara, D.; Kokini, J.L. Bioinformatic approaches for characterizing molecular structure and function of food proteins. *Annu. Rev. Food Sci. Technol.* **2023**, *14*, 203–224. [CrossRef]
26. Patel, B.; Singh, V.; Patel, S. Structural bioinformatics. In *Essentials of Bioinformatics*; Shaik, N.A., Hakeem, K.R., Eds.; Springer: New York, NY, USA, 2019; Volume 1, pp. 169–199.
27. Tandang-Silvas, M.R.G.; Fukuda, T.; Fukuda, C.; Prak, K.; Cabanos, C.; Kimura, A.; Itoh, T.; Mikami, B.; Utsumi, S.; Maruyamaet, N. Conservation and divergence on plant seed 11S globulins based on crystal structures. *Biochim. Biophys. Acta* **2010**, *1804*, 1432–1442. [CrossRef]
28. Santos, M.; Teixeira, T.R.; Santos, F.R.S.; Lima, W.G.; Ferraz, A.C.; Silva, N.L.; Leite, F.J.; Siqueira, J.M.; Luyten, W.; Castro, A.H.F.; et al. *Bauhinia holophylla* (Bong.) Steud. leaves-derived extracts as potent anti-dengue serotype 2. *Nat. Prod. Res.* **2019**, *35*, 2804–2809. [CrossRef] [PubMed]
29. Castro, A.H.F.; Tavares, H.S.; Pereira, S.R.F.; Granjeiro, P.A.; Silva, J.A.; Galdino, A.S. Production and characterization of lectin from *Bauhinia holophylla* (Fabaceae:Cercideae) calli. *Plant Cell Tissue Organ Cult.* **2018**, *134*, 423–432. [CrossRef]
30. Lloyd, G.; Mc Cown, B. Commercially feasible micropropagation of mountain laurel, *Kalmia latifolia*, by use of shoot tip culture. *Int. Plant Propagators' Soc. Proceed.* **1980**, *30*, 421–427.
31. Bala, T.N.N.; Mathur, K.; Kumar, S.; Vyas, M.; Saini, A.; Tomar, B. In vitro callus induction for determination of lectin activity in pea (*Pisum sativum* L.) variety (AP-1). *Rom. Biotechnol. Lett.* **2010**, *15*, 5781–5787.
32. Liener, I.E.; Hill, E.G. The effect of heat treatment on the nutritive value and hemagglutinin activity of soybean oil meal. *J. Nutr.* **1953**, *49*, 609–620. [CrossRef]
33. Inbar, M.; Sachs, L. Structural difference in sites on the surface membrane of normal and transformed cells. *Nature* **1969**, *223*, 710–712. [CrossRef]
34. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids. Symp. Ser.* **1999**, *41*, 95–98. [CrossRef]
35. Wu, A.M.; Wu, J.H.; Liu, J.H.; Singh, T. Recognition profile of *Bauhinia purpurea* agglutinin (BPA). *Life Sci.* **2004**, *74*, 1763–1779. [CrossRef] [PubMed]
36. Souza, J.D.; Silva, M.B.R.; Argolo, A.C.C.; Napoleão, T.H.; Sá, R.A.; Correia, M.T.S.; Paiva, P.M.G.; Silva, M.D.C.; Coelho, L.C.B.B. A new *Bauhinia monandra* galactose-specific lectin purified in milligram quantities from secondary roots with antifungal and termiticidal activities. *Int. Biodeterior. Biodegrad.* **2011**, *65*, 696–702. [CrossRef]
37. Silva, A.L.C.; Horta, A.C.G.; Moreira, R.D.A. Isolation and partial characterization of a lectin from *Bauhinia pentandra* (Bong.) vog. Ex. Steud. *Rev. Bras. Fisiol. Veg.* **2001**, *13*, 262–269. [CrossRef]
38. Chan, Y.S.; Ng, T.B. *Bauhinia variegata* var. variegata lectin: Isolation, characterization, and comparison. *Appl. Biochem. Biotechnol.* **2015**, *175*, 75–84. [CrossRef] [PubMed]
39. Silva, H.C.; Bari, A.U.; Pereira-Jénior, F.N.; Simões, R.C.; Barroso-Neto, I.L.; Nobre, C.B.; Pereira, M.G.; Nascimento, K.S.; Rocha, B.A.M.; Delatorre, P.; et al. Purification and partial characterization of a new pro-inflammatory lectin from *Bauhinia bauhinioides* Mart. (Caesalpinoideae) seeds. *Protein Pept. Lett.* **2011**, *18*, 396–402. [CrossRef] [PubMed]
40. Silva, M.C.C.; Santana, L.A.; Mentele, R.; Ferreira, R.S.; De Miranda, A.; Silva-Lucca, R.A.; Sampaio, M.U.; Correia, M.T.S.; Oliva, M.L.V. Purification, primary structure and potential functions of a novel lectin from *Bauhinia forficata* seeds. *Process Biochem.* **2012**, *47*, 1049–1059. [CrossRef]

41. Silva, H.C.; Pinto, L.D.S.; Teixeira, E.H.; Nascimento, K.S.; Cavada, B.S.; Silva, A.L.C. BUL: A novel lectin from *Bauhinia ungulata* L. seeds with fungistatic and antiproliferative activities. *Process Biochem.* **2014**, *49*, 203–209. [CrossRef]
42. Katoch, R.; Tripathi, A. Research advances and prospects of legume lectins. *J. Biosci.* **2021**, *46*, 104. [CrossRef]
43. Cunningham, A.B.; Brinckmann, J.A.; Harter, D.E.V. From forest to pharmacy: Should we be depressed about a sustainable *Griffonia simplicifolia* (Fabaceae) seed supply chain? *J. Ethnopharmacol.* **2021**, *278*, 114202. [CrossRef]
44. De Coninck, T.; Van Damme, E.J.M. Review: The multiple roles of plant lectins. *Plant Sci.* **2021**, *313*, 111096. [CrossRef] [PubMed]
45. De Simone, G.; Pasquadibisceglie, A.; Proietto, R.; Polticelli, F.; Aime, S.; Op den Camp, H.J.M.; Ascenzi, P. Contaminations in (meta)genome data: An open issue for the scientific community. *IUBMB Life* **2020**, *72*, 698–705. [CrossRef] [PubMed]
46. Loris, R.; Hamelryck, T.; Bouckaert, J.; Wyns, L. Legume lectin structure. *Biochim. Biophys. Acta* **1998**, *1383*, 9–36. [CrossRef] [PubMed]
47. Sharon, N. Lectins: Carbohydrate-specific reagents and biological recognition molecules. *Biol. Chem.* **2007**, *282*, 2753–2764. [CrossRef] [PubMed]
48. Van Holle, S.; De Schutter, K.; Eggermont, L.; Tsaneva, M.; Dang, L.; Van Damme, E.J.M. Comparative study of lectin domains in model species: New insights into evolutionary dynamics. *Int. J. Mol. Sci.* **2017**, *18*, 1136. [CrossRef] [PubMed]
49. Vieira-Silva, S.; Touchona, M.; Abby, S.S.; Rocha, E.P.C. Investment in rapid growth shapes the evolutionary rates of essential proteins. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20030–20035. [CrossRef] [PubMed]
50. Chakraborty, S.; Kahali, B.; Ghosh, T.C. Protein complex forming ability is favored over the features of interacting partners in determining the evolutionary rates of proteins in the yeast protein-protein interaction networks. *BMC Syst. Biol.* **2010**, *4*, 155. [CrossRef] [PubMed]
51. Drummond, A.; Raval, A.; Wilke, C.A. A single determinant dominates the rate of yeast protein evolution. *Mol. Biol. Evol.* **2006**, *23*, 327–337. [CrossRef] [PubMed]
52. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein identification and analysis tools on the expasy server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press: Totowa, New Jersey, USA, 2005; pp. 571–607.
53. Katoch, R.; Tripathi, A.; Hallan, V.; Purohit, R.R. Structure prediction and protein-ligand interaction of lectin from rice bean (*Vigna umbellata*). *Authorea* **2024**. [CrossRef]
54. Streicher, H.; Sharon, N. Recombinant plant lectins and their mutants. *Methods in Enzymol.* **2003**, *363*, 47–77. [CrossRef]
55. Kaushik, S.; Mohanty, D.; Surolia, A. The role of metal ions in substrate recognition and stability of concanavalin A: A molecular dynamics study. *Biophys. J.* **2009**, *96*, 21–34. [CrossRef] [PubMed]
56. Mi, F.; Guana, M.; Hu, C.; Peng, F.; Sun, S.; Wang, X. Application of lectin-based biosensor technology in the detection of foodborne pathogenic bacteria: A review. *Analyst* **2021**, *146*, 429–443. [CrossRef] [PubMed]
57. Kapp, K.; Schrempf, S.; Lemberg, M.K.; Dobberstein, B. Post-targeting functions of signal peptides. In *Madame Curie Bioscience Database [Internet]*; Landes Bioscience: Heidelberg, Germany, 2009. Available online: https://www.ncbi.nlm.nih.gov/books/NBK6322/ (accessed on 23 March 2024).
58. Cheng, P.-N.; Pham, J.D.; Nowick, J.S. The supramolecular chemistry of β-sheets. *J. Am. Chem. Soc.* **2013**, *135*, 5477–5492. [CrossRef] [PubMed]
59. Perczel, A.; Gáspári, Z.; Csizmadia, I.G. Structure and stability of β-pleated sheets. *J. Comput. Chem.* **2005**, *26*, 1155–1168. [CrossRef] [PubMed]
60. Smith, C.K.; Regan, L. Construction and design of β-sheets. *Acc. Chem. Res.* **1997**, *30*, 153–161. [CrossRef]
61. Deller, M.C.; Kong, L.; Rupp, B. Protein stability: A crystallographer's perspective. *Acta Cryst.* **2016**, *72*, 72–95. [CrossRef] [PubMed]
62. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. [CrossRef]
63. Ejigu, G.F.; Jung, J. Review on the computational genome annotation of sequences obtained by next-generation sequencing. *Biology* **2020**, *9*, 295. [CrossRef]
64. Mereghetti, P.; Ganadu, M.L.; Papaleo, E.; Fantucci, P.; Gioia, L. de Validation of protein models by a neural network approach. *BMC Bioinform.* **2008**, *9*, 66. [CrossRef]