SCIENTIA PLENA
www.scientiaplena.org.br
VOL. 20, NUM. 10 2024
doi: 10.14808/sci.plena.2024.100202

In silico characterization, phylogeny and molecular docking of the Bowman-Birk type protease inhibitor in species from Phaseoleae tribe

Caracterização *in silico*, filogenia e docking molecular do inibidor de protease do tipo Bowman-Birk em espécies da tribo Phaseoleae

F. G. Silva[1]*; R. T. Maia[2]; H. J. Jimenez[3]; R. M. Moraes Filho[1]; G. C. Silva[1]

[1]*Departamento de Agronomia/Programa de Pós-Graduação em Melhoramento Genético de Plantas/Universidade Federal Rural de Pernambuco, 56580-000, Recife - Pernambuco, Brazil*
[2]*Centro de Desenvolvimento Sustentável do Semiárido/Programa de Pós-Graduação em Ciências Naturais e Biotecnologia/Universidade Federal de Campina Grande, 58175-000, Cuité - Paraíba, Brazil*
[3]*Secretaria de Educação do Estado de Pernambuco, 50810-900, Recife - Pernambuco, Brazil*

**flavia.gomess@ufrpe.br*
(Recebido em 11 de abril de 2024; aceito em 01 de outubro de 2024)

The subfamily Papilionoideae (Fabaceae) has approximately 32 tribes, the Phaseoleae tribe being the largest, with about 90 genera and 1,600 species, distributed in seven subtribes. This tribe has a great diversity of species of economic importance as a source of human and animal food. They also stand out for the presence of the Bowman-Birk-type protease inhibitor (BBI), which is well recognized for its insecticidal action. Thus, the objectives of this research were to characterize, compare and identify conserved domains in BBI sequences from different species of the tribe Phaseoleae; perform phylogenetic analysis; develop three-dimensional models and molecular docking. 23 BBI sequences from Phaseoleae species were analyzed. Physicochemical parameters, presence of signal peptide cleavage sites, domain identification and estimation of functional effects were performed by ProtParam, TOPCONS, Prodom and SNAP2, respectively. The MEGA 7 program was used to construct the phylogenetic tree using the Maximum Likelihood method. The prediction, evaluation and validation analyzes of the tertiary structure of the models were obtained by the GalaxyTBM, UCSF Chimera, SAVES and ProSA-web servers and the ClusPro-2.0 SPPIDER servers for the molecular docking analysis. For all species, the isoelectric point showed a slightly acidic character. Two functional domains have been identified in the BBI sequences. From the molecular docking analysis it was possible to show interface residues. Cluster analysis was consistent with the most recent phylogenies of the analyzed species.
Keywords: bioinformatics, computational analysis, serine proteases.

A subfamília Papilionoideae (Fabaceae) possui aproximadamente 32 tribos, sendo a tribo Phaseoleae a maior, com cerca de 90 gêneros e 1.600 espécies, distribuídas em sete subtribos. Esta tribo possui uma grande diversidade de espécies de importância econômica como fonte de alimentação humana e animal. Destacam-se também pela presença do inibidor de protease do tipo Bowman-Birk (BBI), que é bastante reconhecido por sua ação inseticida. Assim, os objetivos desta pesquisa foram caracterizar, comparar e identificar domínios conservados em sequências BBI de diferentes espécies da tribo Phaseoleae; realizar análises filogenéticas; desenvolver modelos tridimensionais e docking molecular. Foram analisadas 23 sequências BBI de espécies de Phaseoleae. Parâmetros físico-químicos, presença de sítios de clivagem de peptídeos sinal, identificação de domínios e estimativa de efeitos funcionais foram realizados por ProtParam, TOPCONS, Prodom e SNAP2, respectivamente. O programa MEGA 7 foi utilizado para construir a árvore filogenética utilizando o método de Máxima Verossimilhança. As análises de predição, avaliação e validação da estrutura terciária dos modelos foram obtidas pelos servidores GalaxyTBM, UCSF Chimera, SAVES e ProSA-web e pelos servidores ClusPro-2.0 SPPIDER para a análise de docking molecular. Para todas as espécies o ponto isoelétrico apresentou caráter levemente ácido. Dois domínios funcionais foram identificados nas sequências BBI. A partir da análise de docking molecular foi possível evidenciar resíduos de interface. A análise de agrupamento foi consistente com as filogenias mais recentes das espécies analisadas.
Palavras-chave: bioinformática, análise computacional, serinoproteases.

1. INTRODUCTION

The subfamily Papilionoideae has 32 tribes [1], and the tribe Phaseoleae is the largest with approximately 90 genera and 1,600 species, subdivided into seven subtribes [2]. In addition, this tribe stands out for its great economic importance [3], for encompassing the main grains cultivated for human and animal consumption, such as common beans (*Phaseolus vulgaris*), cowpea (*Vigna unguiculata*) [2, 4], Pigeon pea (*Cajanus cajan*) [2] and soybean (*Glycine max*) [2, 4], belonging to the subtribes Phaseolinae, Cajaninae and Glycininae, respectively.

The members of this tribe have a wide global distribution [5], with the species of the Phaseolinae subtribe distributed from the Old (European, African and Asian continents and the four archipelagos of Macaronesia) to the New World (American continent) [6, 7]. Most species of subtribe Cajaninae are Asian [2], with subtribe Glycininae found in Asia, part of Russia to Australia [8]. In addition, they are characterized by the presence of primary leaves (eophiles) simple and opposite, trifoliolate leaves with asymmetric base, papilionaceous corolla, inflorescences in pseudoracemos, pollen grain covered by a thick layer of endexin, presence of stipels (small stipule), number basic chromosome 11 [9].

Several representatives of legumes (Fabaceae) stand out for containing the Bowman-Birk-type protease inhibitor (BBI) [10], as is the case of species from the Phaseoleae tribe (Papilionoideae). BBI are called proteins with two different inhibitory domains, where generally the first active site is specific for trypsin, and the second inhibits chymotrypsin or elastase [11, 12]. The BBIs were so named because in 1946, Donald E. Bowman [13] isolated the first BBI in soybean (G. max), and it was later characterized in 1963 by Birk et al. (1963)[14]. These BBIs are found abundantly in seeds and tubers of legumes (Fabaceae), Solanaceae, Cucurbitaceae and Poaceae [15].

Several functions of BBIs in plants are mentioned in the literature, the insecticidal action being the most reported. This function is closely related to its activity in the intestinal proteases of insects, which makes it possible to reduce the assimilation of amino acids, which can delay the development of insects that feed on these plants [16]. Certain BBIs may also favor the microbial effect in plants [17], as well as tolerance to abiotic stresses, such as salt [18] and water [19].

Other studies show the roles of BBIs in human health, such as anti-inflammatory properties in intestinal diseases [20, 21] and anticancer properties [22, 23]. It has been shown that the two binding sites of BBI (antitrypsin and antichymotrypsin) are closely linked to the antiproliferative effect on colon cancer cells [24].

With the growth in the number of protein sequences in the databases, consequently the number of functionally related sequences of the same protein group, has been growing exponentially. Thus, it becomes increasingly necessary that the relationships between members of the same protein family be analyzed [25]. In this sense, phylogenetic analyzes can help to understand the relationships of protein groups in different groups of organisms, enabling the understanding of how these proteins are related in different species, and if they evolved from a single common ancestor [26]. To help in this understanding, the prediction of the three-dimensional structure of the protein is also carried out, which allows a better understanding of its molecular structures and functions. From the knowledge of the 3D structure of a protein, molecular docking analysis can also be performed, which consists of predicting the best conformation of a ligand compared to another molecule, building a stable complex [27]. Thus, this analysis plays a significant role in drug development, as well as in the elucidation of important biochemical processes [28].

Increasingly, studies have been carried out to elucidate the structure [12-29], evolutionary and phylogenetic relationships of BBI [30], where phylogenetic results are continually being improved due to the increasing availability of a large amount of of biological data and the emergence of new advances in methods to analyze them [26, 31].

Thus, the objectives of this research were: (1) To characterize, compare and identify conserved domains in Bowman-Birk-type protease inhibitor sequences from different species of the Phaseoleae tribe (Fabaceae - Papilionoideae) available in public databases; (2) Perform a phylogenetic analysis based on the sequences found, to understand their relationships; (3) To

build three-dimensional models of tribe representatives; (4) Perform the moldocular docking analysis.

2. MATERIALS E METHODS

2.1. Recovery of sequences

The BBI sequences were obtained using the BLAST (Basic Local Alignment Search Tool) search algorithm available on the NCBI (National Center for Biotechnology Information) plataform, and were downloaded in FASTA format. BBI sequences from species of the Papilionoideae subfamily were retrieved using the BLASTp tool, which searches for homologous sequences in the database, where a total of 23 sequences were selected (Table 1).

Table 1: Analysis of the primary structure and subcellular location of the Bowman-Birk inhibitor sequences of species from the Phaseoleae tribe (Fabaceae - Papilionoideae). GB-ID: Genbank Identification, SCSP: Signal Peptide Conserved Sites, Nº aa: Number of amino acids, MW: Molecular Weight; pI: Isoelectric Point. GRAVY: Grand Mean, Hydropathicity.

Species	GB-ID	SCPS	Nº aa	MW (kDa)	pI	GRAVY
Phaseolus microcarpus	CAL64060.1	29-30	120	13.19	5.69	-0.222
Phaseolus zimapanensis	CAQ52360.1	29-30	120	13.11	5.61	-0.197
Phaseolus filiformis	CAL69282.1	28-29	120	13.09	5.69	-0.222
Phaseolus grayanus	CAQ52359.1	29-30	120	13.13	5.69	-0.233
Phaseolus hintonii	CAQ52357.1	29-30	119	13.02	5.22	-0.199
Phaseolus oligospermus	CAL51270.1	29-30	120	13.07	6.08	-0.166
Phaseolus augusti	CAL51269.1	28-29	120	13.03	5.61	-0.187
Phaseolus glabellus	CAL69238.1	29-30	120	13.04	5.18	-0.154
Phaseolus lunatus	CAL51268.1	28-29	120	13.09	5.61	-0.194
Phaseolus coccineus subsp. polyanthus	CAQ52820.1	29-30	120	13.05	5.86	-0.232
Phaseolus vulgaris	CAM34246.1	28-29	120	13.10	6.22	-0.232
Phaseolus coccineus subsp. coccineus	CAL69277.1	28-29	117	12.82	5.69	-0.195
Phaseolus costaricensis	CAL69280.1	28-29	120	13.04	5.44	-0.182
Phaseolus maculatus	CAL51271.1	28-29	120	13.03	5.86	-0.137
Phaseolus acutifolius var. latifolius	CAL49456.1	29-30	118	12.92	5.79	-0.204
Vigna marina	ABD97867.1	21-22	112	12.16	6.22	-0.072
Vigna mungo	AAK97766.1	21-22	103	11.31	6.80	-0.085
Vigna subterranea	AHY03234.1	29-30	121	13.15	6.50	-0.214
Glycine soja	XP_028181512.1	27-28	114	12.49	5.44	-0.108
Glycine max	NP_001236353.2	27-28	114	12.48	5.42	-0.108
Glycine microphylla	AAO89510.1	-	93	10.23	5.41	-0.567
Rhynchosia sublobata	ALA09300.2	22-23	112	12.41	5.44	-0.304
Cajanus cajan	KYP42282.1	21-22	101	11.22	6.00	-0.155

2.2. Sequence analysis

The ProtParam server (http://web.expasy.org/protparam) analyzed the physicochemical parameters of BBI present in species from the tribe Phaseoleae (subfamily Papilionoideae). The presence of signal peptide cleavage sites was analyzed using the TOPCONS server (http://topcons.cbr.su.se/pred). To identify the functional domains of the protein, the Prodom server (http://prodom.prabi.fr/prodom/) was used for its classification and ontology. The

functional effects caused by mutations in amino acid sequences were estimated using the SNAP2 server (https://rostlab.org/services/snap2web/).

2.3. Alignment and Phylogenetic Analysis

BBI protein sequences were aligned using the ClustalW algorithm and the phylogenetic tree was produced in MEGA 7.0 software [32]. The phylogenetic tree was developed using the Maximum Likelihood (ML) method, with statistical support calculated by the bootstrap method (BS) with 1000 replicates.

2.4. Modeling, evaluation and validation of tertiary structures

The tertiary structure prediction of the BBI was performed by the GalaxyTBM server (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=TBM) in Template Based Modeling mode. The 3D structure visualization was built using the UCSF Chimera package [33]. The server SAVES v6.0 (https://saves.mbi.ucla.edu/results?job=1294919&p=procheck) was used to verify the quality of the model by analyzing the Ramachandran graph. The Z-score calculation was performed on the ProSA-web interactive server (https://prosa.services.come.sbg.ac.at/prosa.php), to identify errors in three-dimensional structures [34].

Trypsin (GenBank: AHL46496.1) and chymotrypsin (NCBI Reference Sequence: XP_049706047.) protease sequences from Helicoverpa armigera were also modeled and validated by the tools used for BBI modeling. The GHECOM server (https://pdbj.org/ghecom/) was used to predict the active sites of the models.

2.5. Molecular docking

The models obtained for BBI were submitted to molecular docking with models of protease structures (trypsin and chymotrypsin). For the docking simulations, the ClusPro-2.0 server (https://cluspro.bu.edu/login.php) was used in the default settings, with the BBIs treated as binding molecules and the proteases as receptor molecules. The complexes with the best ranking (most clustered) were selected for visual analysis and the SPPIDER- Solvent accessibility based Protein-Protein Interface Identification and Recognition server (https://sppider.cchmc.org/) was used to identify residues in the interface of protein-protein interaction.

3. RESULTS AND DISCUSSION

Based on analysis of the primary structure and subcellular location of BBI sequences from species of the Phaseoleae tribe (Table 1), the conserved sites of signal peptides ranged from 22 to 30. The signal peptide is a short, transient sequence that has the purpose of function to direct the proteins for secretion or for transfer to their characteristic organelles and for further processing [35]. After directing the protein to the correct transport pathway, it is cleaved by signal peptidases [36].

There was a variation in the amino acid size of BBI from 93 aa (*Glycine microphylla*) to 121 aa (*Vigna subterranea*). The small size of BBI sequences, usually between 60 and 90, differentiates it from other families of protease inhibitors [37]. Tiessen et al. (2012) [38], point out that the composition of amino acids is influenced by the size of the protein, thus the amino acids of small proteins tend to be different from larger proteins [38]. For the molecular weight, a variation of 10.23kDa (*G. microphylla*) to 13.19kDa (*P. microcarpus*) was obtained, this information can be applied to isolate proteins, from chromatography and electrophoresis techniques [39].

The isoelectric point (pI) attested to a slightly acid character, which varied from 5.18 (*P. glabellus*) to 6.80 (*Vigna mungo*). In the literature, pI values of different proteins are

observed, ranging from highly acidic to very alkaline values from 4.0 to 12.0 [40]. This variation is associated with the difference between the amino acid residues of different BBI [41]. Kiraga et al. [42] also showed that the pI can be widely varied, depending on the insertions and deletions between protein orthologs and the ecology of the organism.

This information is very significant for knowing the solubility and subcellular localization of proteins, which indicates that it is an important resource for predicting their net charge at a given pH and understanding the interactions between proteins, proteins and membranes (phospholipids), as well as for indicate the existence of protein isoforms [43]. Knowing the pI of a protein is an important parameter for many biochemical and proteomic analyses, such as 2D-PAGE gel electrophoresis [44], capillary isoelectric focusing [45], liquid chromatography coupled with mass spectrometry [46] and crystallography of X-rays [47].

Mean hydropathicity ranged from -0.072 (*V. marina*) to -0.567 (*G. microphylla*). This parameter is used to demonstrate the protein hydrophobicity value, which is obtained from the calculation of the sum of the hydropathy values of all amino acids, divided by the size of the sequence [48], with positive averages for hydrophobic proteins and negative for hydrophobic proteins. hydrophilic proteins. Thus, the average hydropathicity values indicated that the BBI protein is hydrophilic for all species. Average hydropathicity has important applications, such as helping to understand the 3D structure of proteins, their evolutionary relationship, and the location of analogous and distant proteins [49].

According to the multiple alignment of sequences (Figure 1), it was possible to verify regions of high similarity, where most of the positions showed conservation between the BBI sequences of the analyzed species.

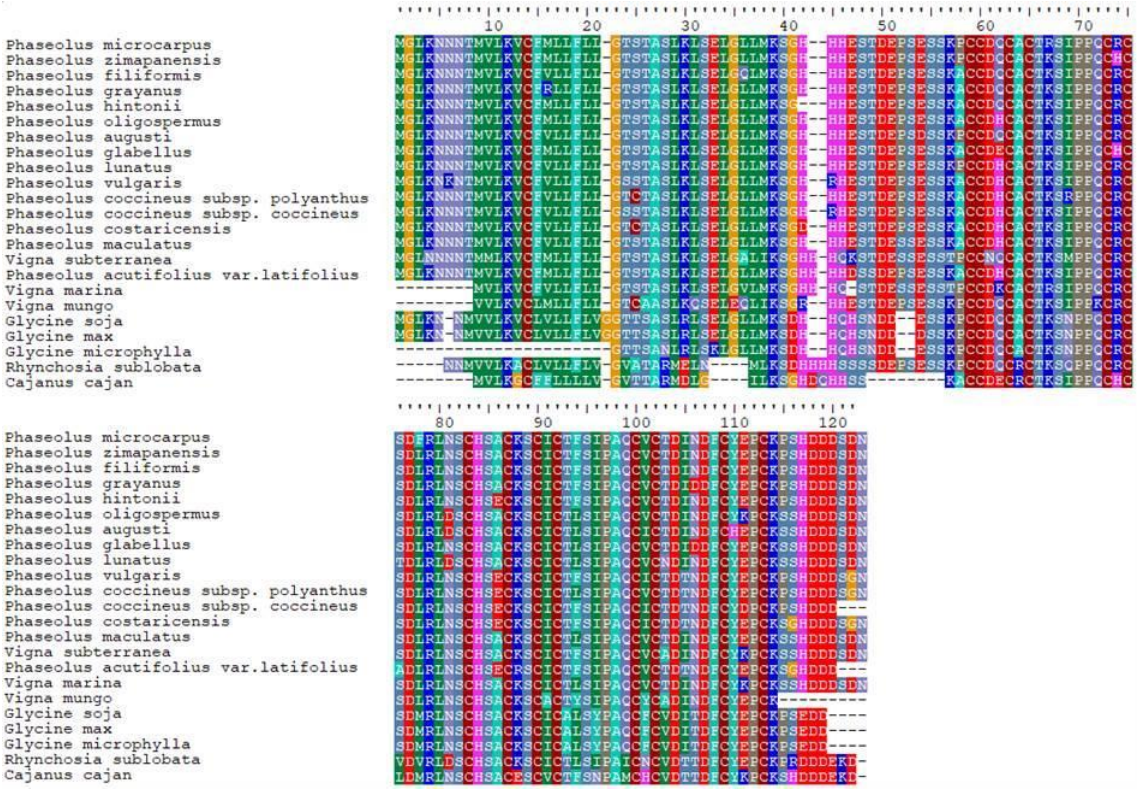

Figure 1: Bowman-Birk inhibitor sequences alignment. Sequences were aligned by ClustalW, where identical and similar residues are displayed in the same color.

Clemente and Arques (2014) [10], when performing the amino acid sequence alignment of the main BBI isoinhibitors from soybeans and other representative species of legumes, also showed that there is a high sequence homology between the BBI isoforms.

For the functional characterization of BBI and prediction of the functional effects of mutations, four species were selected (Figure 2).

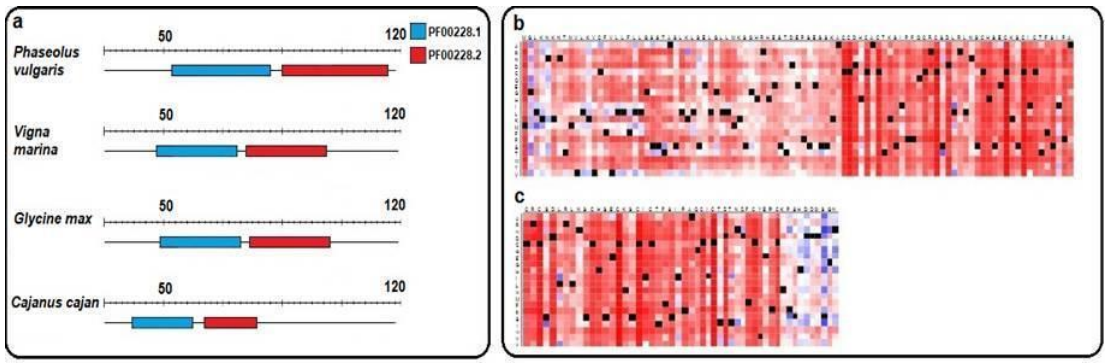

Figure 2: Functional domains observation through the ProDom server (a). Prediction of functional effects of amino acid mutations by SNAP2 server for the PF00228.2 and PF00228.1 domains in Phaseolus vulgaris (b) and Cajanus cajan (c) species.

Two functional domains were identified: PF00228.1 and PF00228.2 in the BBI amino acid sequences. Since these two functional domains are found in all observed sequences, they are associated with inhibitor activity functions of two serine proteases of different specificity [50]. Generally, the first site inhibits trypsin, and the second chymotrypsin [11].

From the action of these domains, several defensive functions of BBIs in plants are reported, the insecticidal action being the most studied, which are associated with its activity in intestinal proteases, which reduces the assimilation of amino acids, which delays the development of insects that become infected. nourish these plants [16]. Kim et al. (2009) [17] also showed that some BBIs can inhibit proteases produced by pathogens, which enables their antimicrobial effect. In addition to these functions, these inhibitors are associated with tolerance to abiotic, saline [18] and water stress [19].

In this way, the use of BBIs can point genes for the genetic engineering of economically important plants, allowing the development of transgenic plants resistant to insects [18], as well as to pathogens and most abiotic stresses. This would make it possible to reduce labor costs and the use of agricultural pesticides against insect and pathogen attack, to provide a more sustainable environment.

In addition to these functions, studies also demonstrate that BBIs have anti-inflammatory properties [51] as well as anti-cancer properties [51]. The anticancer properties are linked to the two binding sites of BBI (antitrypsin and antichymotrypsin activities), providing an antiproliferative effect on cancer cells (mainly colon) [24].

The SNAP2 server showed that the functional domains PF00228.1 and PF00228.2 are the protein regions most sensitive to mutations. This server provided a map with the presence of different shades with possible replacement in each position of these proteins. The red hue signals strong signals for mutation, the white hue indicates low effect, and the blue hue indicates neutrality (Figure 2). The occurrence of a greater number of sites sensitive to mutations in the PF00228.1 and PF00228.2 domains can be explained by their conservation and incidence in the sequences analyzed here. Other works, based on analyzes to predict the functional effects of mutations in amino acids by the SNAP2 server and functional domains of putrescine N-methyltransferase in Solanaceae species [52] and mannose-specific lectin in *Allium* species [53], also showed a higher incidence of sites sensitive to domain mutations.

The phylogenetic analysis based on the BBI resulted in a monophyletic tree (Figure 3), where the formation of three main clades comprising three subtribes of the Phaseoleae tribe (Fabaceae - Papilionoideae) was observed. The largest clade was composed by the species of the subtribe Phaseolinae (BS=86%), represented by the genera *Phaseolus* and *Vigna*, the second clade was formed by the species representative of the subtribe Glycininae (*Glycine microphylla*,

G. soja and *G. max*) (BS= 100%) and the third clade integrated by the species of the Cajaninae subtribe (*Rhynchosia sublobata* and *C. cajan*) (BS=86%).

In the clade of the subtribe Phaseoleolinae, the formation of three clades within the genus Phaseolus was observed, where the first group (BS=20%) comprises the species *P. microcarpus*, *P. zimapanensis*, *P. hintonii*, *P. filiformis* and *P. grayanus*, the second group (BS=52%) with *P. acutifolius var. latifolius*, *P. coccineus subsp. polyanthus*, *P. costaricensis*, *P. vulgaris* and *P. coccineus subsp. coccineus* and the third (BS=7%) formed by the species *P. glabellus*, *P. maculatus*, *P. oligospermus*, *P. augusti* and *P. lunatus*. In the clade of the genus *Vigna*, it was possible to observe that the species *V. mungo* presented an isolated group in the phylogenetic tree (BS=86%), which is also close to the species *V. subterranea* and *V. marina* (BS=56%).

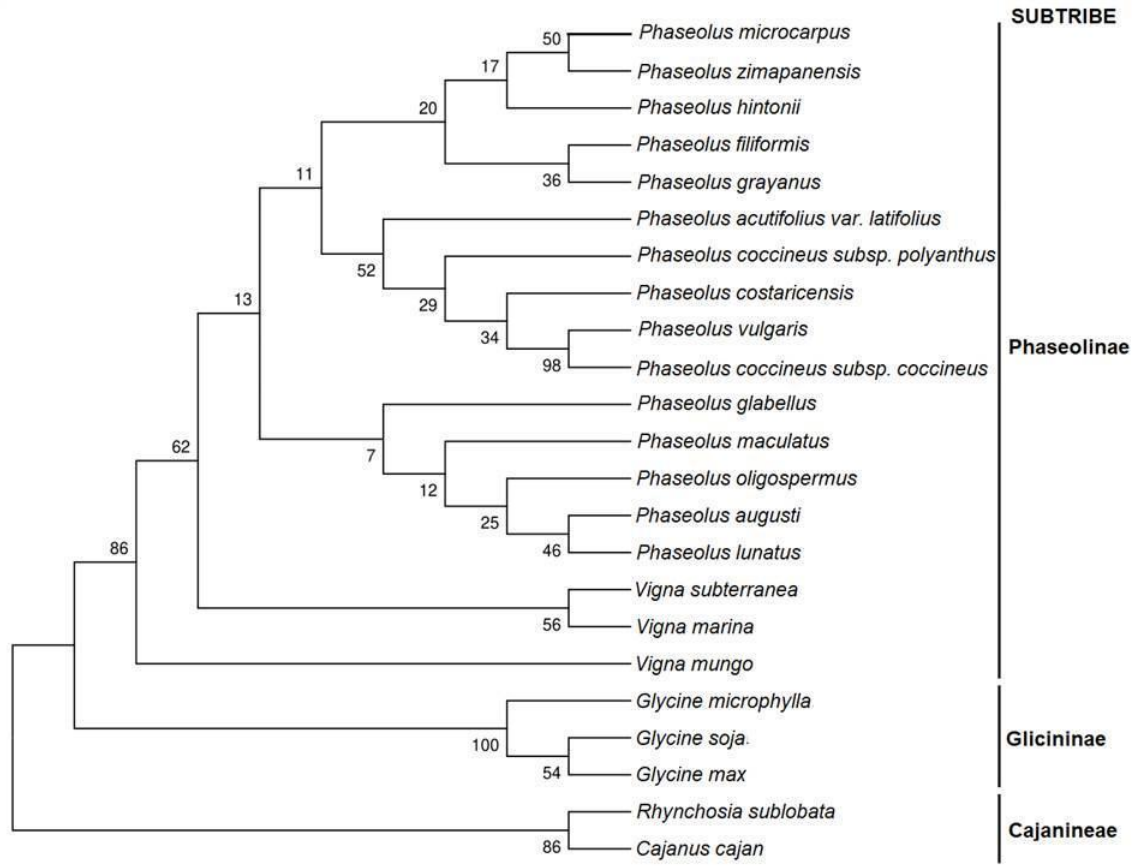

Figure 3: Phylogenetic tree of species from the Phaseoleae tribe (Fabaceae - Papilionoideae), generated based on Bowman-Birk inhibitor sequences by the Maximum Likelihood method.

Species of the genus *Phaseolus* are all endemic to the New World, with species of the first subclade found mainly from Mexico to California (*P. filiformis*) [6]. In the second subclade, species are distributed from southeastern Canada south to the eastern US and from the southern US to southeastern California, through Mexico and Central America, as well as into the Andean region of South America [6]. In the third subclade, the species *P. glabellus* occurs from southern Nuevo Leon and Tamaulipas to central Chiapas (Mexico), *P. oligospermus*, *P. augusti* and *P. lunatus* from Mexico to Central America and parts of South America, and *P. maculatus* from the eastern US to eastern Texas and southern Mexico [6]. In this sense, the geographic proximity of these species supports the formation of the *Phaseolus* clade.

Jin et al. (2019) [54], based on the plastid genome of species from the tribe Desmodieae and Phaseoleae, showed that common bean (*P. vulgaris*) and cowpea (*V. unguiculata*) were reconstructed as a monophyletic group. In other studies, from chloroplast DNA, it was possible to verify that species of the New World genus *Vigna* were more closely related to other New World genera of Phaseolinae (such as *Phaseolus*) than to Old World *Vigna* [6, 7], as can be seen

in *V. mungo* which was isolated from the other species of *Vigna*. Species from almost all American genera have a “right-handed” floral asymmetry as a synapomorphy [7]. These species are all climbers and located in tropical savannah forests [55], although it is possible to locate some subgroups that diversified in montane tropical forests, as is the case of the genus *Phaseolus* [6].

In the clade Glycininae (*G. microphylla*, *G. soybean* and *G. max*), the grouping of *G. max* and *G. soybean* species is probably related to their genetic proximity, since these species belong to the subgenus soybean, which includes *G. max* and *G. soja* (wild relative of commercial soybean), which are native to East Asia and parts of Russia. *G. microphylla*, on the other hand, belongs to the subgenus *Glycine*, which includes about 30 perennial species, whose center of diversity is found in Australia, distributed in diverse habitats, from the desert to temperate and subtropical climates [8].

Phylogenetic analyzes from plastid gene sequences of phaseoloid legume species, allowed the visualization of a strongly supported group (BS=98% and BS=85%) that included the subtribes Phaseolinae, Glycininae and Cajaninae [56-57], where the monophyletic grouping of *Rhynchosia* and *Cajanus* was evidenced [56], corroborating the results observed here.

In the subtribe Cajaninae, studies by Cândido et al. (2020)[58], based on the analysis of the ITS nuclear and plastid regions *rpl32* and *trnQ*, also supported the grouping of *Cajanus* with *R. volubilis*. Similar characteristics are shared by these two genera, such as the habit that can vary from subshrub to herbaceous, trifoliate to unifoliate leaves (in some *Rhynchosia*), with the greatest difference observed in the fruit, which can have two to three seeds (or more) in *Rhynchosia* and *Cajanus*, respectively. The species *R. sublobata* generally has trifoliate leaves, a climbing habit, yellow flowers streaked with red and calyx lobes passing through the corolla [59]. Another shared characteristic is the geographical distribution, where *Rhynchosia* species are Asian, as well as most species of the genus *Cajanus* [2].

The low support relationships presented in the phylogenetic tree can be justified by the inefficiency of the marker (BBI) to separate or organize genetically very close species.

Phylogenetic analyzes based on protein sequences have been widely used to clarify the function of proteins within a group of organisms [59]. Phylogenetic reconstructions may explain questions about how proteins are related in different species, and whether they may have evolved from a common ancestor [60]. BBI sequence-based phylogeny demonstrated similarity with other Papilionoideae phylogenies from plastid genes [61-64].

To predict the tertiary structure of the BBI the species *P. vulgaris* and *C. cajan*, representatives of the Phaseoleae tribe (Fabaceae - Papilionoideae), were selected (Figure 4).

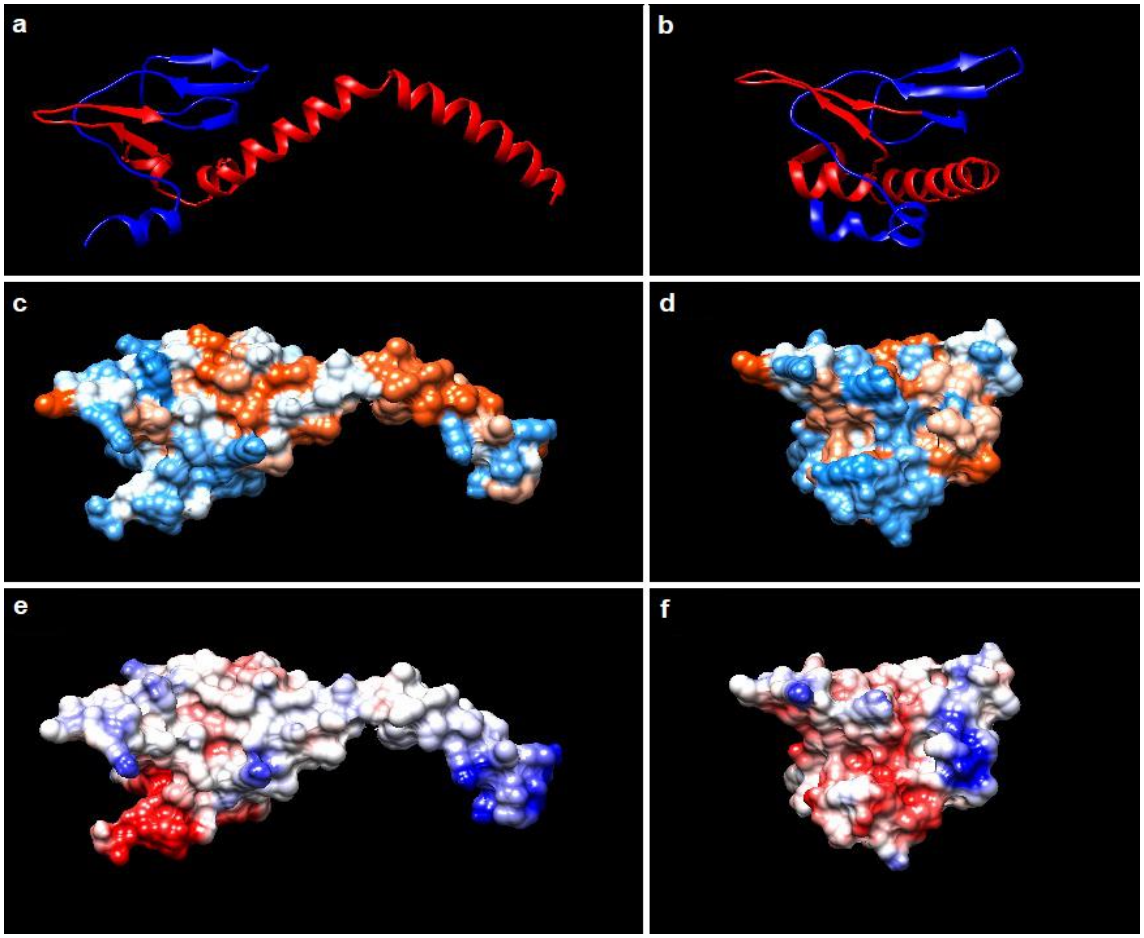

Figure 4: Three-dimensional structure predicted by the GalaxyTBM server for the Bowman-Birk inhibitor of the species Phaseolus vulgaris (a) and Cajanus cajan (d). Hydrophobicity represented as a color gradient, with blue being the most hydrophilic and orange-red being the most hydrophobic of the species P. vulgaris (b) and C. cajan (e). Electrostatic surface represented as a color gradient, from the most negatively charged (red) to the most positively charged (blue) of the species P. vulgaris (c) and C. cajan (f).

The PF00228.1 and PF00228.2 domains exhibit a similar conformation, so these domains probably have the same functionality in the two analyzed species, as was also observed in the tertiary structure of BBI from *V. unguiculata* seeds [29]. The two domains present a conserved conformation, justified by their presence in the two analyzed species, also observed by other authors in their 3D models for other types of proteins [52,53]. Chen et al. (1992) [65], states that the three-dimensional structure of BBI is stabilized, mainly by the presence of seven disulfide bridges, together with a little help from the hydrophobic core of the molecule.

In the trypsin and chymotrypsin models of the species *Helicoverpa armigera*, the presence of beta and alpha-helix sheets for both enzymes can be observed (Figure 5).

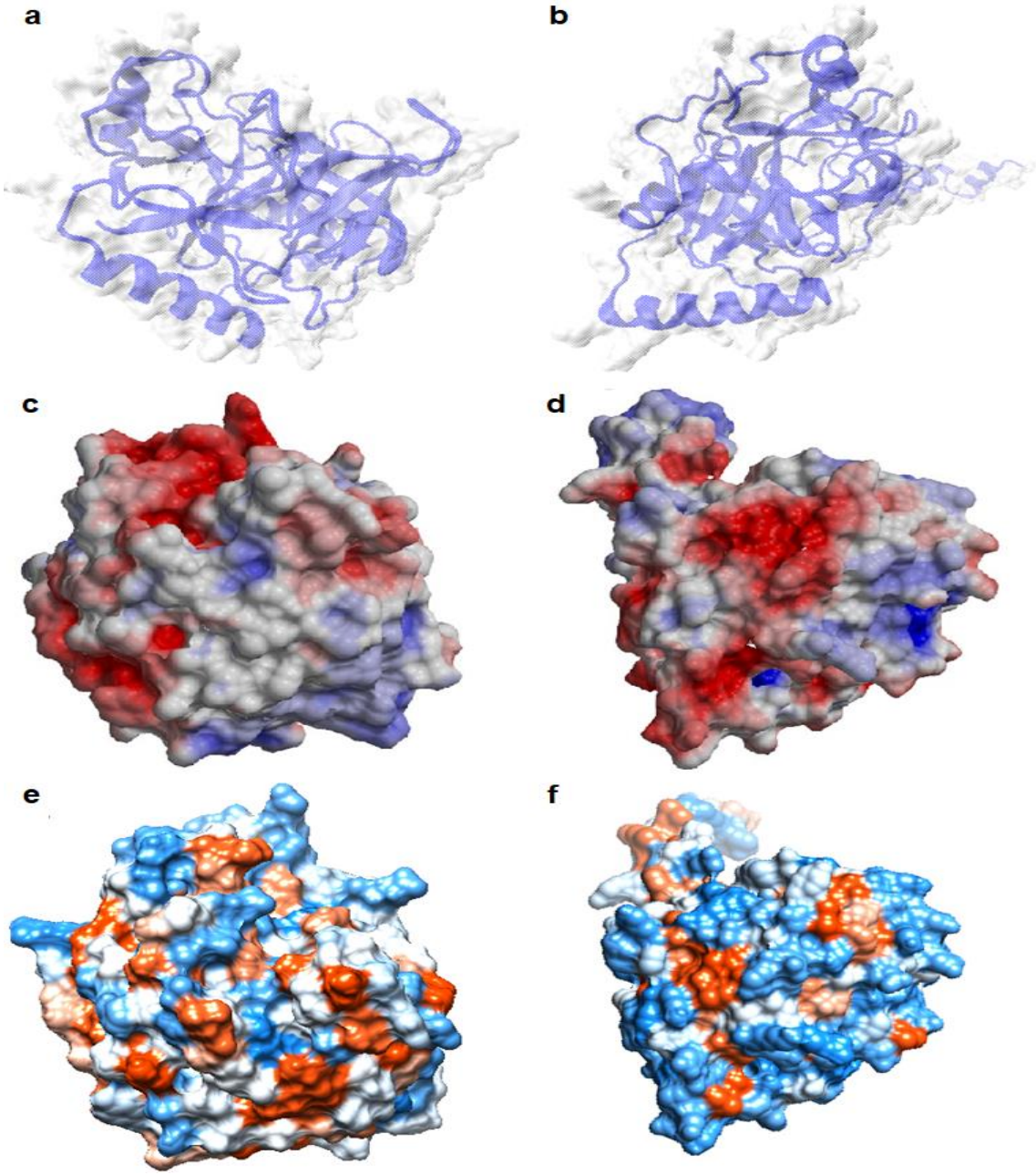

Figure 5: Three-dimensional structure predicted by the GalaxyTBM server for the proteases trypsin (a) and chymotrypsin (b) of the species Helicoverpa armigera. Electrostatic surface represented as a color gradient, from the most negatively charged (red) to the most positively charged (blue) of trypsin (c) and chymotrypsin (d). Hydrophobicity represented as a color gradient, with blue being the most hydrophilic and orange-red being the most hydrophobic of trypsin (e) and chymotrypsin (f).

The 3D models were validated by Ramachandran plot and Z-score. The designation of these models is of paramount importance, since the determination of the three-dimensional structures are generated by the coding sequences, which allow, therefore, to evaluate with more precision the biological effects. Since, the prediction of 3D structures of proteins usually results in practical applications of relevance in pharmaceutical industry [66].

The refined models showed 93.5 and 96.9% of amino acid residues in the most favoured region by the analysis of the Ramachandran graph, for the species *P. vulgaris* and *C. cajan* respectively (Figure 6). For trypsin and chymiotrypsin models the Ramanchandran plot pointed respectively 90.6% and 91% of residues in most favoured region (Figure 7).

Through analysis on the ProSA server, Z-score values of -3.89, -3.26, -6.11 and 6.78 were observed for *C. cajan* and *P. vulgaris*, trypsin and chymiotrypsin respectively (Figure 6 and Figure 7).

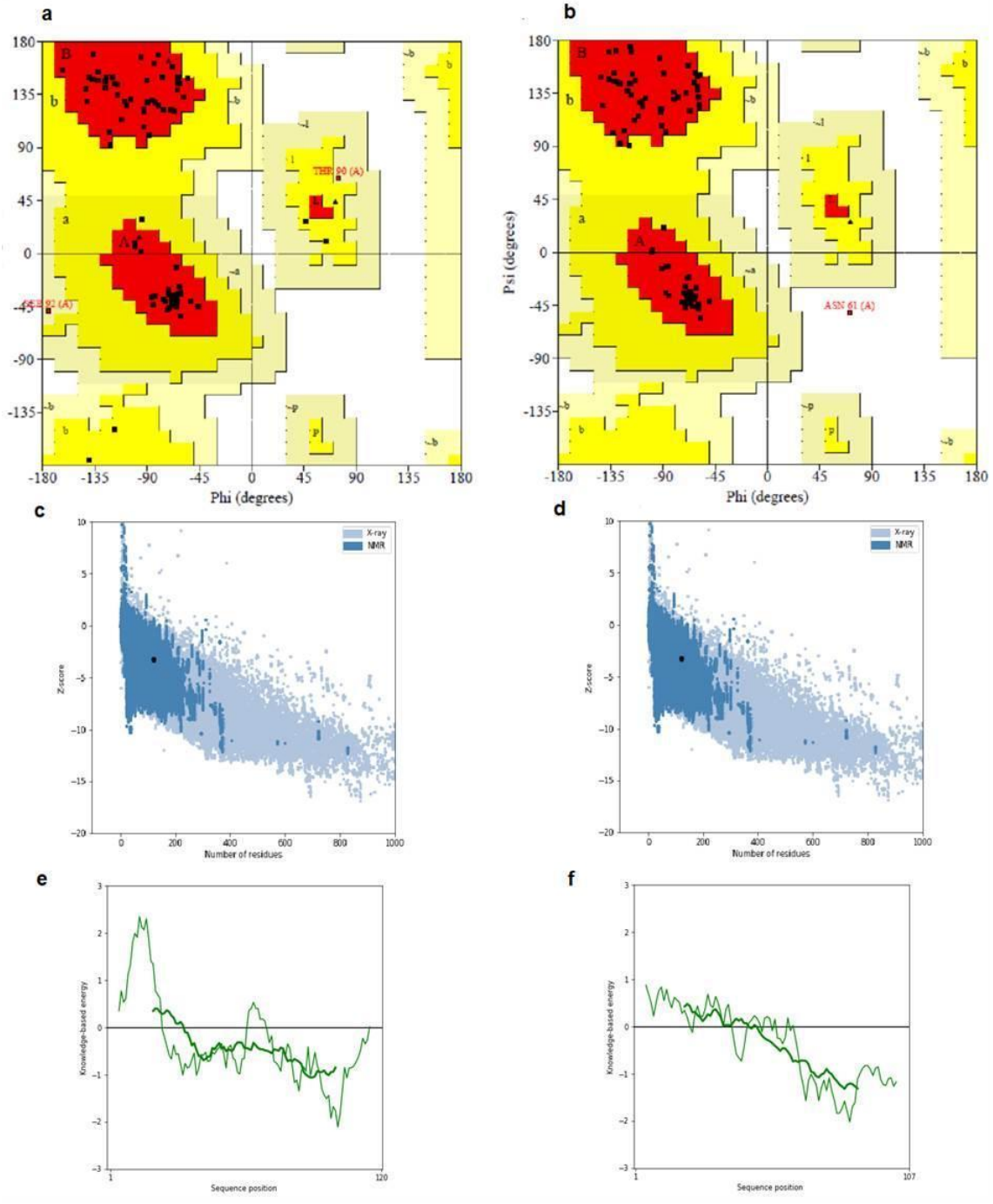

Figure 6: Ramachandran plot for BBI of Phaseolus vulgaris (a) and Cajanus cajan (b). Z-score value (black point) by ProSA-web for P. vulgaris (c) and C. cajan (d). Use of ProSA-web for Bowman-Birk inhibitor showing energy plot of native protein structure residue scores for P. vulgaris (e) and C. cajan (f).

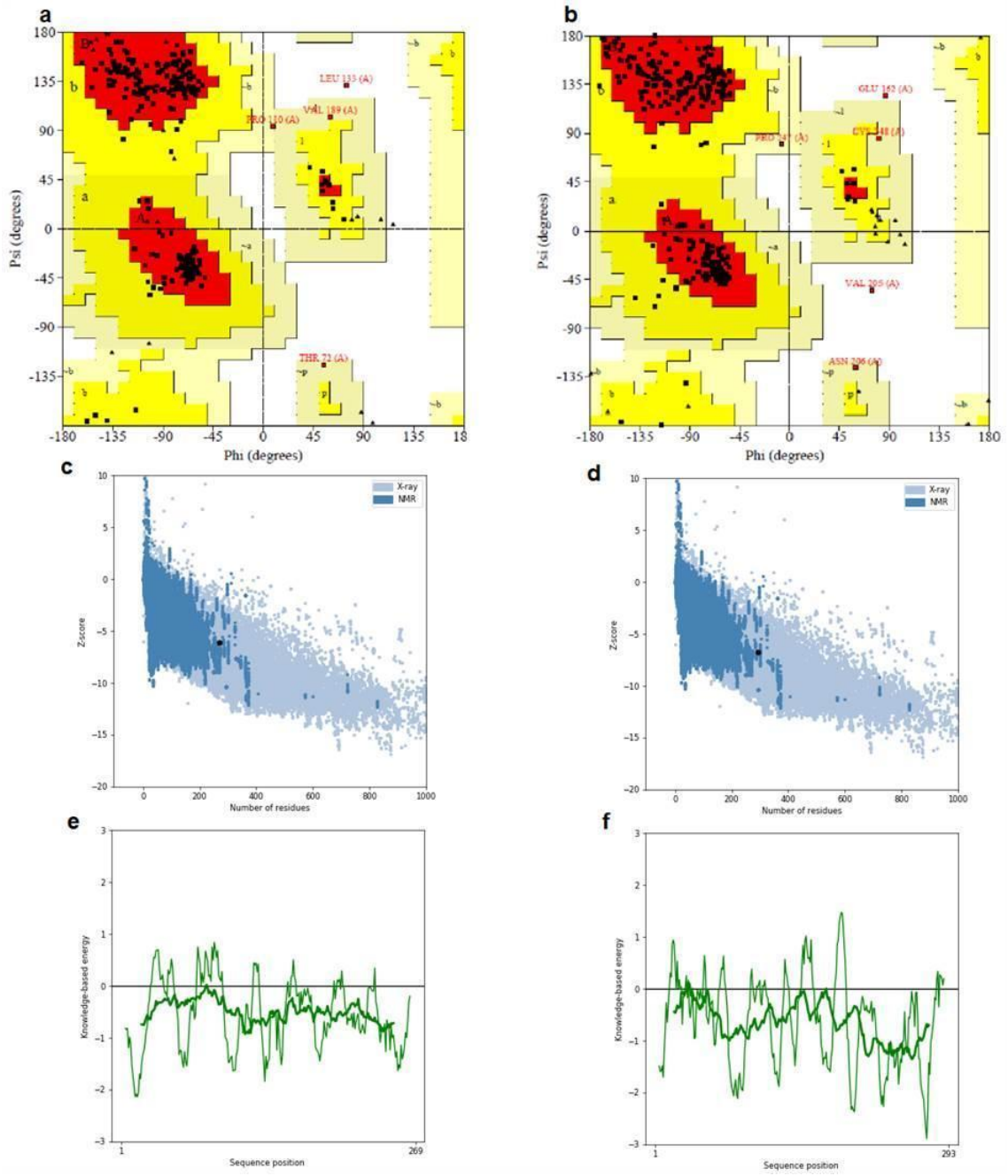

Figure 7: Ramachandran plot for trypsin (a) and chymotrypsin (b) from Helicoverpa armigera. Z-score value (black dot) by ProSA-web for trypsin (c) and chymotrypsin (d). Use of ProSA-web for Bowman-Birk inhibitor showing energy plot of native protein structure residue scores for trypsin (e) and chymotrypsin (f).

The quality of the model is obtained from the result of the Z-score of protein structures resolved as references [67] This parameter makes it possible to understand to what extent the quality of the model deviates from the real crystalline structure. In this sense, a Z-score greater than zero indicates the model is the most ideal [68]. That way, the values of Z-scores indicate that the structures are erroneous when this value is outside a characteristic range for native proteins [34]. Thus, it can be confirmed that the Z-scores obtained for the BBI structure of *P. vulgaris* and *C. cajan* (-3.26 and -3.89) are within the range of scores typically identified for native proteins of similar sizes.

The functionality of proteins is associated with their 3D configuration, which in turn is determined by their amino acid sequence [52]. Although sequence analysis is complex,

computational methods have become increasingly used in protein prediction. Thus, comparative or homology modeling is the most used, because it effectively generates good results, being able to predict the real structure of a protein from a known amino acid sequence, with a degree of homology greater than 30% [69].

In all docking simulations (Table 2), with the exception of docking between *C. cajan* BBI and trypsin, cluster 0 was the most representative in populations of complexed structures. All results pointed to an affinity between proteases and BBIs. The highlight was the result between *C. cajan* BBI with chymotrypsin, which presented the cluster with the highest population (275 structures) and the lowest energy scores (-941.3/-1215.2).

Table 2: Scores of the best represented clusters in each docking simulation.

Cluster	Members	Representative	Score
P. vulgaris BBI/trypsin (Cluster 0)	102	Center	-775.6
		Lowest Energy	-845.4
P. vulgaris BBI/chymiotrypsin (Cluster 0)	118	Center	-931.0
		Lowest Energy	-1003.6
C. cajan BBI/trypsin (Cluster 1)	76	Center	-760.1
		Lowest Energy	-891.6
C. cajan BBI/chymiotrypsin (Cluster 0)	275	Center	-941.3
		Lowest Energy	-1215.2

When analyzing the interaction residues (Table 3), it can be noted that none of the trypsin interface residues converge with those indicated in cluster 1 of the GHECOM graph. For chymotrypsin, 10 of the 27 residues that interact with the *P. vulgaris BBI* participate in the supposed active site of this protease, however, none of the 23 residues of the *P. vulgaris* BBI inhibitor belong to the predicted active site.

In the trypsin-BBI/*C. cajan*, it is observed that only one residue (F9) from 22 listed in the interaction belongs to the catalytic site of BBI, while none of the 24 aminoacids listed for trypsin are present in cluster 1 of GHECOM.

On the chymotrypsin-BBI/C interaction *C. cajan* the presence of 10 of the 26 active site residues involved in the interaction with the inhibitor was observed. Concomitantly, it was observed that 9 of the 26 aminoacids of BBI interaction are from the active site of this molecule. These results, combined with the docking simulation, point to this interaction as the most promising of the data obtained, since in addition to lower energy and a larger clustered population in docking, there was a direct interaction of the active sites of both enzymes. In Figure 8, one can observe the interaction regions of the best complex obtained in this study, the BBI/*C* complex. *cajan*-chymotrypsin.

Table 3: List of residues (amino acids) that interface in the protein-protein interaction of the complexes obtained in molecular docking.

Protein-protein interaction	**Residues (amino acids)**
Trypsin residues that interact with *P. vulgaris* BBI (1262 Ä2)	A5 Y6 I8 L9 A10 L12 A13 V14 A16 A17 A18 P19 I23 V24 G25 G26 D27 H28 T30 I31 E33 Y134 P137 N139 F140 P143 A161 L163 K167 R207 V267 A268 L269 **(33)**
P. vulgaris BBI residues that interact with trypsin (1320 Ä2)	K4 T8 V10 L11 C14 F15 L18 F19 L21 G22 T25 A26 K29 L30 L33 M37 H41 R42 H43 T46 D47 L77 N78 P94 A95 **(25)**
Chymotrypsin residues that interact with *P. vulgaris* BBI (1037 Ä2)	E33 F35 T36 A37 S38 I68 W69 W72 H73 F74 H90 R94 D129 Y130 I132 Y177 P178 G179 R180 L181 K203 S204 V205 H22 S228 S243 P2475 **(27)**
P. vulgaris BBI residues that interact with chymotrypsin (1209 Ä2)	M1 L3 K4 K6 N7 M9 V10 L11 V13 C14 F15 L17 L18 F19 L20 L21 G22 S24 T25 K29 N78 P94 A95 **(23)**
Trypsin residues that interact with *C. cajan* BBI (897 Ä2)	M1 A2 A5 Y6 L9 L12 Y134 L135 P137 N139 F140 D144 S217 T246 N247 L250 E251 I262 R263 A265 P266 V267 A268 L269 **(24)**
C. cajan BBI residues that interact with trypsin (872 Ä2)	F9 L13 V16 R20 L23 K27 N61 S62 C63 S65 A66 C67 E68 S69 C70 V71 C72 F74 S75 A78 H81 V83 **(22)**
Chymotrypsin residues that interact with *C. cajan* BBI (1204 Ä2)	S32 E33 F35 T36 A37 I68 W69 W72 F74 R94 L95 K96 Y99 D129 Y130 I132 Y177 P178 G179 R180 L181 S204 V205 H225 P247 R250 **(26)**
C. cajan BBI residues that interact with chymotrypsin (1218 Ä2)	M2 V3 G6 C7 F9 L10 L11 L13 V14 T17 T18 R20 M21 L23 K27 L60 N61 C72 F74 S75 N76 P77 K94 S95 D98 D99 **(26)**

In red, residues that are part of the catalytic site of the enzyme as predicted by the GHECOM server.

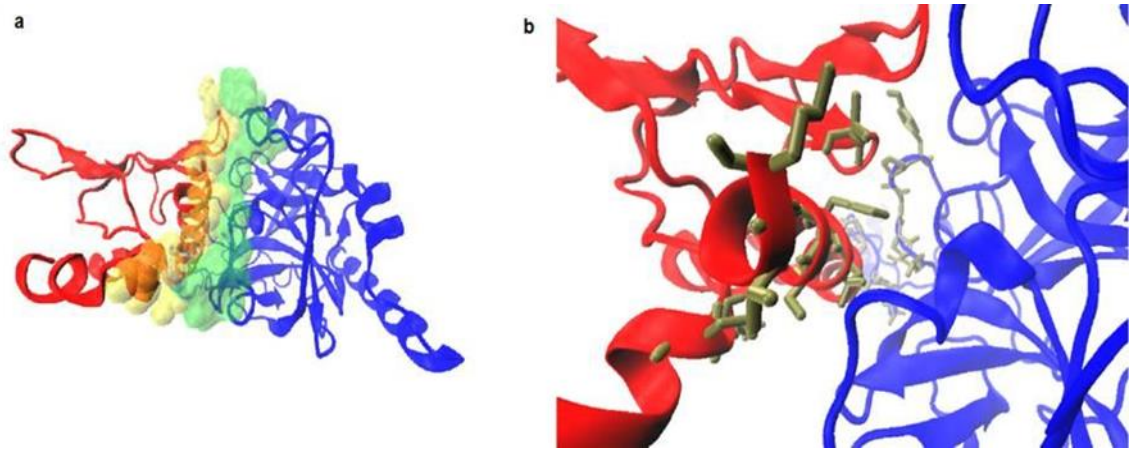

Figure 8: In red, the Cajanus cajan BBI inhibitor. In blue, chymotrypsin. In a, the regions with interface residues of each enzyme are represented in yellow (BBI) and green (chymotrypsin) surfaces. In b, represented in sticks, the interface amino acids that also participate in the active site of the respective enzymes.

Interactions occur primarily between chymotrypsin loop regions and a BBI inhibitor alpha-helix region. This alpha helix has active site residues, and likely plays a crucial role in inhibiting the protease chymotrypsin.

Would the BBIs from *C. cajan* and *P. vulgaris* be protease inhibitors from *Helicoverpa armigera*? Potentially, yes. Low energy scores indicate affinity between the enzymes in all complexes. Although trypsin did not present a direct interaction, the possibility of inhibition of this protease by allosteric interactions cannot be ruled out, considering the large amount of amino acids that interact with this protease. Molecular dynamics simulations in future studies may elucidate this possibility of allosteric interaction [70]. Chymotrypsin was interacted directly in regions belonging to the active site by the two BBI inhibitors studied here, highlighting the interaction with BBI from *C. cajan*. Despite the interesting and suggestive results for the inhibition of proteases, it must be taken into account that the theoretical and predictive approach of molecular docking does not in itself constitute definitive evidence, thus demanding future experimental studies in vitro and in vivo.

The supposed anti-protease activity of BBI inhibitors deserves future studies and attention, as these molecules may have great potential for pest control. The use of inhibitors at safe concentrations can provide excellent targeted and sustainable biological control. Studies with transgenics to increase the expression of insect protease inhibitor genes can be extremely useful for obtaining pest resistant varieties. Another possibility is the rational design of bioinspired peptides that mimic the inhibitory action of BBIs for the future production of biodegradable pesticides.

4. CONCLUSION

The present study allowed a better understanding of the functions of BBI in the inhibitory activity of the two serine proteases. Associated with this, the grouping obtained from the BBI protein sequences confirms the recent taxonomic classifications of the species of the tribe Phaseoleae (Fabaceae - Papilionoideae).

The use of in silico tools proved to be valid and reliable in the construction of 3D models, in the generation of stable models regarding their energy level and within the standards, which is capable of being considered an analogous representation of the real structure of the already established protein, which provides the analysis of their tertiary structures as well as their molecular functions. It was also possible to show interface residues from molecular docking analysis.

5. ACKNOWLEDGMENT

This research was supported by the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES). The author F. G. Silva would like to thank CAPES for the doctoral scholarship granted.

6. REFERENCES

1. Polhill RM, Raven PH, Stirton CH. Evolution and systematics of the Leguminosae. In Polhill RM, Raven PH, editors. Advances in legume systematics. Richmond (US): Royal Botanic Gardens; 1981. p. 1-26.
2. Schrire BD. Tribe Phaseoleae. In: Lewis G, Schrire BD, Mackinder B, Lock M, editors. Legumes of the world. Richmond (US): Royal Botanic Gardens; 2005. p. 393-431.
3. Teixeira JD, Ferreira JJS, Silva JS. A tribo Phaseoleae s.l. (Leguminosae-Papilionoideae) no município de Caetité, Estado da Bahia, Brasil. Hoehnea. 2021 Apr;48(1):1-30. doi: 10.1590/2236-8906-39/2020
4. Barreto KL, Fernandes MF, Queiroz LP. Flora of Bahia: Leguminosae – Centrosema (Papilionoideae: Phaseoleae). Sitientibus Ser Ci Biol. 2020 Nov;20:1-23. doi: 10.13102/scb5280
5. Tozzi AMGA. Papilionoideae. In: Tozzi AMGA, Melhem TS, Forero E, Fortuna-Perez AP, Wanderley MGL, Martins SE, et al., editors. São Paulo: Instituto de Botânica; 2016. p. 167-397.

6. Delgado-Salinas A, Bibler R, Lavin M. Phylogeny of the genus *Phaseolus* (Fabaceae): A recent diversification in an ancient landscape. Syst Bot. 2006 Oct;31(4):779-91. doi: 10.1600/036364406779695960
7. Delgado-Salinas A, Thulin M, Pasquet R, Weeden N, Lavin M. *Vigna* (Fabaceae) sensu lato: the names and identities of the american segregate genera. Am J Bot. 2011 Oct;98(10):1694-715. doi: 10.3732/ajb.1100069
8. Sherman-Broyles S, Bombarely A, Powell AF, Doyle JL, Egan An, Coate JE, et al. The wild side of a major crop: soybean's perennial cousins from down under. Am J Bot. 2014 Oct;101(10):1651-65. doi: 10.3732/ajb.1400121
9. Espert SM, Sede SM, Ruiz LK, Fortunato RH, Poggio L. New chromosome reports in the subtribes Diocleinae and Glycininae (Phaseoleae: Papilionoideae: Fabaceae). Bot J Linn Soc. 2008 Sep;158(2):336-41. doi: 10.1111/j.1095-8339.2008.00864.x
10. Clemente A, Arques MC. Bowman-Birk inhibitors from legumes as colorectal chemopreventive agentes. World J Gastroenterol. 2014 Aug;20(30):10305-15. doi: 10.3748/wjg.v20.i30.10305
11. Lioi L, Galasso I, Daminati MG, Piergiovanni AR. Inhibitory properties and binding loop polymorphism in Bowman-Birk inhibitors from *Phaseolus* species. Genet Resour Crop Evol. 2010 Oct;57:533-54. doi: 10.1007/s10722-009-9491-6
12. Qi RF, Song ZW, Chi CW. Structural features and molecular evolution of Bowman-Birk protease inhibitors and their potential application. Acta Biochim Biophys Sin. 2005 May;7(5):283-92. doi: 10.1111/j.1745-7270.2005.00048.x
13. Bowman DE. Differentiation of soy bean anti-tryptic factors. Proc Soc Exp Biol Med. 1946 Dec;63(3):547-50. doi: 10.3181/00379727-63-15668
14. Birk Y, Gertler A, Khalef S. A pure trypsin inhibitor from soya beans. Biochem J. 1963 May;87(2):281-4. doi: 10.1042/bj087028i
15. Connors BJ, Laun NP, Maynard CA, Powell WA. Molecular characterization of a gene encoding a cystatin expressed in the stems of American chestnut (*Castanea dentata*). Planta. 2002 Jul;215(3):510-4. doi: 10.1007/s00425-002-0782-9
16. Deepika R, Nath AK. Studies on trypsin inhibitor in bean (*Phaseolus vulgaris* L) cultivars of himalayan region. J Plant Biochem Biotechnol. 2010 Dec;19:103-5. doi: 10.1007/BF03323445
17. Kim JY, Park SC, Hwang I, Cheong H, Nah JW, Hahm KS, et al. Protease inhibitors from plants with antimicrobial activity. Int J Mol Sci. 2009 Jun 23;10(6):2860-72. doi: 10.3390/ijms10062860
18. Shan L, Li C, Chen F, Zhao S, Xia GA. Bowman-Birk type protease inhibitor is involved in the tolerance to salt stress in wheat. Plant Cell Environ. 2008 Aug;31(8):1128-37. doi: 10.1111/j.1365-3040.2008.01825.x
19. Dramé KN, Passaquet C, Repellin A, Zuily-Fodil Y. Cloning characterization and differential expression. of a Bowman-Birk inhibitor during progressive water deficit and subsequent recovery in peanut (*Arachis hypogaea*) leaves. J Plant Physiol. 2013 Jan;15;170(2):225-9. doi: 10.1016/j.jplph.2012.09.005
20. Utrilla MP, Peinado MJ, Ruiz R, Rodriguez-Nogales A, Algieri F, Rodriguez-Cabezas ME, et al. Pea (*Pisum sativum* L.) seed albumin extracts show anti-inflammatory e ect in the DSS model of mouse colitis. Mol Nutr Food Res. 2015 Apr;59(4):807-19. doi: 10.1002/mnfr.201400630
21. Juritsch AF, Moreau R. Role of soybean-derived bioactive compounds in inflammatory bowel disease. Nutr Rev. 2018 Aug;76(8):618-38. doi: 10.1093/nutrit/nuy021
22. Fereidunian A, Sadeghalvad M, Oscoie MO, Mostafaie A. Soybean bowman-birk protease inhibitor (BBI): Identification of the mechanisms of BBI suppressive e ect on growth of two adenocarcinoma cell lines: AGS and HT29. Arch Med Res. 2014 Aug;45(6):455-61. doi: 10.1016/j.arcmed.2014.07.001
23. Oliás R, Becerra-Rodríguez C, Soliz-Rueda JR, Moreno FJ, Delgado-Andrade C, Clemente A. Glycation a ects dierently the main soybean Bowman-Birk isoinhibitors, IBB1 and IBBD2, altering their antiproliferative properties against HT29 colon cancer cells. Food Funct. 2019 Sep;1;10(9):6193-202. doi: 10.1039/c9fo01421g
24. Clemente A, Moreno FJ, Marín-Manzano MDC, Jiménez E, Domoney C. The cytotoxic e ect of Bowman-Birk isoinhibitors, IBB1 and IBBD2, from soybean (*Glycine max*) on HT29 human colorectal cancer cells is related to their intrinsic ability to inhibit serine proteases. Mol Nutr Food Res. 2010 Mar;54(3):396-405. doi: 10.1002/mnfr.200900122
25. Watson JD, Laskowski RA, Thornton JM. Predicting protein function from sequence and structural data. Curr Opin Struc Biol. 2005 Jun;15(3):275-84. doi: 10.1016/j.sbi.2005.04.003
26. Kasap S, Benkrid K. High performance phylogenetic analysis with maximum parsimony on reconfigurable hardware. IEEE Transactions on VLSI Systems. 2010 Jun;19(5):796-808. doi: 10.1109/TVLSI.2009.2039588

27. Lengauer T, Rarey M. Computational methods for biomolecular docking. Curr Opin Struct Biol. 1996 Jun;6(3):402-6. doi: 10.1016/s0959-440x(96)80061-3
28. Mostashari-Rad T, Arian R, Mehridehnavi A, Fassihi A, Ghasemi F. Study of CXCR4 chemokine receptor inhibitors using QSPR andmolecular docking methodologies. J Comput Chem. 2019 Aug;178(4):1950018. doi: 10.1142/S0219633619500184
29. Barbosa JARG, Silva LP, Teles RCL, Esteves GF, Azevedo RB, Ventura MM, et al. Crystal structure of the bowman-birk inhibitor from *Vigna unguiculata* seeds in complex with b-Trypsin at 1.55 A˚ resolution and its structural properties in association with proteinases. Biophys J. 2007 Mar;92(5):1638-50. doi: 10.1529/biophysj.106.090555
30. James AM, Jayasena AS, Zhang J, Berkowitz O, Secco D, Knott GJ, et al. Evidence for ancient origins of Bowman-Birk Inhibitors from *Selaginella moellendorffii*. Plant Cell. 2017 Mar;29(3):461-73. doi: 10.1105/tpc.16.00831
31. Andrade LF, Nahum LA, Avelar LGA. Eukaryotic protein kinases (ePKs) of the helminth parasite *Schistosoma mansoni*. BMC Genomics. 2011 May;12:215. doi: 10.1186/1471-2164-12-215
32. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016 Jul;33(7):1870-4. doi: 10.1093/molbev/msw054
33. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605-12. doi: 10.1002/jcc.20084
34. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in threedimensional structures of proteins. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W407-10. doi: 10.1093/nar/gkm290
35. Choo KH, Tan TW, Ranganathan S. SPdb – a signal peptide database. BMC Bioinformatics. 2005 Oct;6:249. doi: 10.1186/1471-2105-6-249
36. Peng C, Shi C, Cao X, Li Y, Liu F, Lu F. Factors influencing recombinant protein secretion efficiency in gram-positive bacteria: signal peptide and beyond. Front Bioeng Biotechnol. 2019 Jun;7:139. doi: 10.3389/fbioe.2019.00139
37. Jaulent AM, Leatherbarrow RJ. Design, synthesis and analysis of novel bicyclic and bifunctional protease inhibitors. Protein Eng Des Sel. 2004 Sep;17(9):681-7. doi: 10.1093/protein/gzh077
38. Tiessen A, Pérez-Rodríguez P, Delaye-Arredondo LJ. Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes. BMC Res Notes. 2012 Feb;5:85. doi: 10.1186/1756-0500-5-85
39. Wu YT, Huang WY, Lin TC, Sheu SJ. Determination of moutan tannins by high-performance liquid chromatography and capillary electrophoresis. J Sep Science. 2003 Dec;26(18):1629-34. doi: 10.1002/jssc.200301587
40. Tokmakov A, Kurotani A, Sato K-I. Protein pI and intracellular localization. Front Mol Biosci. 2021 Nov;8:775736. doi: 10.3389/fmolb.2021.775736
41. Losso JN. The biochemical and functional food properties of the Bowman-Birk Inhibitor. Nutr Food Sci. 2008 Jan;48(1):94-118. doi: 10.1080/10408390601177589
42. Kiraga J, Mackiewicz P, Mackiewicz D, Kowalczuk M, Biecek P, Polak N, et al. The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. BMC Genomics. 2007 Jun;8:163. doi: 10.1186/1471-2164-8-163
43. Stekhoven SFM, Gorissen MH, Flik G. The isoelectric point, a key to understanding a variety of biochemical problems: a minireview. Fish Physiol Biochem. 2008 Aug;34:1-8. doi: 10.1007/s10695-007-9145-6
44. Bezerra Júnior RQ, Martins GR, Barroso IC, Marinho RC, Aguiar TDF, Teixeira MFS. Eletroforese Bidimensional e espectrometria de massa como ferramentas proteômica aplicadas à definição de marcadores proteicos associados à eficiência reprodutiva de caprinos. Acta Vet Bras. 2013 Aug;7(2):100-12.
45. Righetti PG, Castagna A, Herbert B, Reymond F, Rossier JS. Prefractionation techniques in proteome analysis. Proteomics. Aug;3(8):1397-407. doi: 10.1002/pmic.200300472
46. Heller M, Ye M, Michel PE, Morier P, Stalder D, Jünger MA, et al. Added value for tandem mass spectrometry shotgun proteomics data validation through isoelectric focusing of peptides. J Proteome Res. 2005 Nov-Dec;4(6):2273-82. doi: 10.1021/pr050193v
47. Kirkwood J, Hargreaves D, O'Keefe S, Wilson J. Using isoelectric point to determine the pH for initial protein crystallization trials. Bioinformatics. 2015 May;31:1444-51. doi: 10.1093/bioinformatics/btv011
48. Chang KY, Yang JR. Analysis and prediction of highly effective antiviral peptides based on random forests. PLoS ONE. 2013 Aug5;8(8):e70166. doi: 10.1371/journal.pone.0070166

49. Lolkema JS, Slotboom DJ. Estimation of structural similarity of membrane proteins by hydropathy profile alignment. Mol Membr Biol. 1998 Jan-Mar;15(1):33-42. doi: 10.3109/09687689809027516
50. Hellinger R, Gruber CW. Peptide-based protease inhibitors from plants. Drug Discov Today. 2019 Sep;24(9):1877-89. doi: 10.1016/j.drudis.2019.05.026
51. Gitlin-Domagalska A, Maciejewska A, D¸ebowski D. Bowman-Birk inhibitors: Insights into family of multifunctional proteins and peptides with potential therapeutical applications. Pharmaceuticals. 2020 Nov 25;13(12):421. doi: 10.3390/ph13120421
52. Gama BRA, Moraes Filho RM, Montarroyos AVV, Martins LSS. In silico characterization of putrescine N-methyltransferase in Solanaceae species. Genet Mol Res. 2022;21:gmr19076. doi: 10.4238/gmr19076
53. Jimenez HJ, Martins LSS, Carvalho R, Montarroyos AVV, Moraes Filho RM. In silico characterization and phylogenetic analysis of a mannose-specific lectin in *Allium* species. Genet Mol Res. 2019;18:gmr18187. doi: 10.4238/gmr18187
54. Jin DP, Choi IS, Choi BH. Plastid genome evolution in tribe Desmodieae (Fabaceae: Papilionoideae). PLoS ONE. 2019 Jun;14:e0218743. doi: 10.1371/journal.pone.0218743
55. Maxted N, Mabuza-Dlamini P, Moss H, Padulosi S, Jarvis A, Guarino L, editors. An ecogeographic study African vigna. Systematic and ecogeographic studies on crop genepools. IPGRI; 2004. Available from: https://hdl.handle.net/10568/105017.
56. Kajita T, Ohashi H, Tateishi Y, Bailey CD, Doyle JJ. *rbcL* and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and Allies. Syst Bot. 2001 Jan;26(3):515-36. doi: 10.1043/0363-6445-26.3.515
57. Stefanovic´ S, Pfeil BE, Palmer JD, Doyle J. Relationships among phaseoloid legumes based on sequences from eight chloroplast regions. Syst Bot. 2009 Mar;34(1):115-28. doi: 10.1600/036364409787602221
58. Cândido ES, Vatanparast M, Vargas W, Bezerra LMPA, Lewis GP, Mansano VF, et al. Molecular phylogenetic insights into the evolution of *Eriosema* (Fabaceae): a recent tropical savanna-adapted genus. Bot J Linn Soc. 2020 Dec;194(4):439-59. doi: 10.1093/botlinnean/boaa059
59. Vargas W, Fortuna-Perez AP, Lewis GP, Piva TC, Vatanparast M, Machado SR. Ultrastructure and secretion of glandular trichomes in species of subtribe Cajaninae Benth. (Fabaceae, Phaseoleae). Protoplasma. 2019 Mar;256(2):431-45. doi: 10.1007/s00709-018-1307-0
60. Van Holle S, De Schutter K, Eggermont L, Tsaneva M. Comparative study of lectin domains in model species: new insights into evolutionary dynamics. Int J Mol Sci. 2017 May;18(6):1136. doi: 10.3390/ijms18061136
61. Cardoso D, Pennington RT, Queiroz LP, Boatwright JS, van Wyk BE, Wojciechowski MF, et al. Reconstructing the deep-branching relationships of the papilionoid legumes. S African J Bot. 2013 Nov;89:58-75. doi: 10.1016/j.sajb.2013.05.001
62. LPWG (Legume Phylogeny Working Group). Legume phylogeny and classification in the 21st century: progress, prospects and lessons for other species-rich clades. Taxon. 2013 Dec;62(2):217-48. doi: 10.12705/622.8
63. LPWG (Legume Phylogeny Working Group). A new subfamily classification of the Fabaceae based on a taxonomically comprehensive phylogeny. Taxon. 2017 Feb;66(1):44-77. doi: 10.5061/dryad.61pd6
64. Zhao Y, Zhang R, Jiang KW, Qi J, Hu Y, Guo J, et al. Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen fixing symbiosis in Fabaceae. Mol Plant. 2021 May;14(5):748-73. doi: 10.1016/j.molp.2021.02.006
65. Chen P, Rose P, Love R, Wei CH, Wang BC. Reactive sites of an anticarcinogenic Bowman–Birk proteinase inhibitor are similar to Other trypsin inhibitors. J Biol Chem. 1992 Jan;267(3):1990-4. doi: 10.2210/pdb1pi2/pdb
66. Schmidt T, Bergner A, Schwede T. Modelling three-dimensional protein structures for applications in drug design. Drug Discov Today. 2014 Jul;19(7):890-7. doi: 10.1016/j.drudis.2013.10.027
67. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics. 2011 Feb;27(3):343-50. doi: 10.1093/bioinformatics/btq662
68. Binbay FA, Rathod DC, George AAP, Imhof D. Quality assessment of selected protein structures derived from homology modeling and AlphaFold. Pharmaceuticals. 2023;16(12):1662. doi: 10.3390/ph16121662
69. Gromiha MM, Nagarajan R, Selvaraj S. Protein Structural Bioinformatics: An Overview. In Ranganathan S, Gribskov M, Nakai K, Schönbach C, editors. Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics. Oxford (UK): Academic Press; 2018. p. 445-9.

70. Bowerman S, Wereszczynski J. Detecting allosteric networks using molecular dynamics simulation. Methods Enzymol. 2016;578:429-47. doi: 10.1016/bs.mie.2016.05.027