International Journal of Advanced Biochemistry Research

ISSN Print: 2617-4693
ISSN Online: 2617-4707
NAAS Rating (2025): 5.29
IJABR 2025; SP-9(12): 600-610
www.biochemjournal.com
Received: 12-09-2025
Accepted: 16-10-2025

Aruna R
Department of Molecular Biology and Biotechnology, S.V. Agricultural College, ANGRAU, Andhra Pradesh, India

Srividhya A
Associate Professor & HOD, Seed Science and Technology, APGC, ANGRAU, Andhra Pradesh, India

Lal Ahamed M
Professor & HOD, Genetics and Plant Breeding, Agricultural College, ANGRAU, Andhra Pradesh, India

Devaki K
Scientist Entomology, Regional Agricultural Research Station, ANGRAU, Andhra Pradesh, India

Latha P
Professor & HOD, Crop Physiology, Regional Agricultural Research Station, ANGRAU, Andhra Pradesh, India

Corresponding Author:
Aruna R
Department of Molecular Biology and Biotechnology, S.V. Agricultural College, ANGRAU, Andhra Pradesh, India

In silico analysis of structural, evolutionary, and physicochemical properties of bowman-Birk protease inhibitors across legume species

Aruna R, Srividhya A, Lal Ahamed M, Devaki K and Latha P

DOI: https://www.doi.org/10.33545/26174693.2025.v9.i12Sh.6571

Abstract
Bowman-Birk inhibitors (BBIs) are small, cysteine-rich serine protease inhibitors widespread in legumes, characterized by their remarkable stability and dual inhibitory activity against trypsin and chymotrypsin. In this study, representative BBI sequences were retrieved from UniProtKB for nine legume species, *Arachis hypogaea, Arachis duranensis, Cicer arietinum, Cajanuscajan, Glycine max, Vigna radiata, Phaseolus angularis, Pisum sativum*and *Medicago sativa*. Multiple sequence alignment (MSA), domain architecture mapping, phylogenetic analysis, and physicochemical characterization were conducted to investigate structural conservation, reactive loop diversity, and evolutionary adaptation. Results revealed highly conserved cysteine frameworks forming canonical disulfide linkages essential for β-trefoil fold stability, while reactive loop motifs (CTKS/CTRS) exhibited lineage-specific divergence correlating with protease target specificity. Phylogenetic and pairwise identity analyses delineated clear evolutionary clusters corresponding to legume subfamilies. Biochemical property analysis demonstrated that dual-loop inhibitors possess higher molecular weight and moderate stability, whereas single-loop forms are smaller and more stable, reflecting functional specialization. These findings underscore the modular evolution of BBIs from ancestral single-domain to double-headed forms, balancing structural rigidity and adaptive protease inhibition.

Keywords: Bowman-Birk inhibitor (BBI), multiple sequence alignment, domain architecture, evolutionary diversification, thermostability, protease defense

Introduction
Bowman-Birk inhibitors (BBIs) constitute one of the most studied families of serine protease inhibitors in plants, first discovered in soybeans by Bowman (1946) [3] and later characterized by Birk (1961) [1] . These inhibitors are abundant in legume seeds and play key roles in plant defence through their dual inhibition of trypsin-and chymotrypsin-like proteases. Structurally, BBIs are small (6-15 kDa), cysteine-rich proteins stabilized by multiple disulfide bridges, conferring exceptional thermal and proteolytic stability (Birk, 1961; Laskowski & Kato, 1980) [1, 6]. The canonical BBI consists of two reactive loops, each functioning as an independent inhibitory domain, linked by a short peptide segment. This dual-loop "double-headed" structure enables simultaneous inhibition of two serine proteases. This study integrates multiple sequence alignment (MSA), domain analysis, and in silico biochemical characterization of BBIs from nine legumes to elucidate structural conservation, evolutionary diversification and functional adaptation within this important protein family.

Material and Methods
Sequence Collection of Bowman-Birk Protease Inhibitors
Dataset of Bowman-Birk type proteinase inhibitor (BBIs) of multiple legumes has been obtained from UniProtKB (Table1) for multiple sequence alignment (MSA) and comparative motif/domain analysis. Selected species represent diverse legume subfamilies and evolutionary lineages, providing a comprehensive framework to assess structural and functional divergence within the BBI family. The metadata summarizes the identifiers for the ten BBI precursor sequences analyzed in this study (Table 1). Each entry includes the UniProt/TrEMBL accession ID, organism of origin, NCBI/UniProt taxonomic identifier (OX), and gene name (GN).

This standardized information ensures traceability of all sequences used in downstream analyses such as multiple sequence alignment, motif identification, and structural evaluation. This structured metadata table forms the foundational reference for the comparative study of reactive-loop diversity, cysteine frameworks, and domain architectures across the BBI family. By consolidating organismal, taxonomic, and gene-level annotations, the table supports reproducibility, facilitates cross-database validation, and provides clarity for readers and researchers accessing the original sequences for further functional or evolutionary analyses.

Table 1: Bowman-Birk Inhibitor (BBI) genes considered across legume species in the study

Accession ID	Organism	Taxon ID (OX)	Gene Name (GN)
A0A151TBJ9	*Cajanus cajan*	3821	KK1_019025
A0A1S2XQ25	*Cicer arietinum*	3827	LOC101502102
A0A6B9VB66	*Arachis hypogaea*	3818	DS421_19g657400
A0A6P4BNC4	*Arachis duranensis*	130453	LOC107466013
P01055 (IBB1)	*Glycine max*	3847	—
P01061 (IBB2)	*Phaseolus angularis*	3914	—
P01062	*Vigna radiata var. radiata*	3916	—
P56679	*Pisum sativum*	3888	—
Q40329	*Medicago sativa*	3879	ATI_18
Q4VVG2	*Vigna unguiculata*	3917	—

Multiple Sequence Alignment (MSA) and Cysteine Framework Analysis

Multiple sequence alignments were performed using Clustal Omega v1.2.4 with default gap penalties and the Gonnet substitution matrix. Alignments were visualized and manually curated in Jalview v2.11. The alignment focused on identifying: Conserved cysteine residues and disulfide framework patterns; Reactive loop motifs containing trypsin/chymotrypsin-binding residues; Insertions/deletions influencing domain configuration. Cysteine positions were mapped and shaded according to occupancy frequency (≥40% threshold). Conserved disulfide connectivity (C-X(2)-C-X(4-5)-C-X(9-10)-C-X(1)-C-X(2)-C-X(10-12)-C) was verified against the canonical I12.001 BBI family reference from InterPro (IPR000864). Sequence logos depicting cysteine conservation and reactive loop motifs were generated using WebLogo 3.0.

Identification of Reactive Loop Motifs and Specificity Determinants

Reactive loop regions corresponding to trypsin and chymotrypsin inhibitory sites were identified manually from the MSA and verified through motif scanning with MEME Suite v5.5.5. Conserved motifs viz. TTCXCTKSI (trypsin-binding) and CTRSIPPC/CTRTSPPC (chymotrypsin-binding) were extracted and compared among species. The P1 residue (Lys/Arg for trypsin; Leu/Phe for chymotrypsin) was annotated following the numbering convention of Bode & Huber (1992) [3]. Sequence alignments were cross-validated with Prosite entries (PS00283 and PS00284) to confirm motif classification. Structural annotations of the reactive loops were projected onto the soybean BBI crystal structure (PDB ID: 2BIJ) using PyMOL v2.5 for positional mapping.

Domain Architecture Analysis

Signal Peptide Prediction

Signal peptides were predicted using SignalP v6.0 (DTU Health Tech, Denmark) under the eukaryotic model. Only sequences with a signal peptide probability ≥0.90 were considered secretory. Hydrophobicity of the signal region was confirmed via ProtScale (Kyte-Doolittle scale).

Domain and Motif Annotation

Domain architectures were identified using InterProScan v5.64, Pfam v35.0, and SMART databases. Functional regions including the signal peptide, disulfide-rich core, reactive loops, and C-terminal tail were color-coded and mapped using TBtools v2.1. The β-trefoil fold was annotated based on consensus cysteine spacing and fold topology.

Visualization

Annotated domain diagrams were generated in Illustrator for Biological Sequences (IBS 1.0) and TBtools. Structural modularity and reactive loop localization were visually aligned across species to compare dual versus single-domain configurations.

Phylogenetic Analysis

Aligned BBI sequences were subjected to phylogenetic analysis using both PHYLIP v3.69 and MEGA X. The Neighbor-Joining (NJ) method was applied with Poisson correction, and 1,000 bootstrap replicates were performed to assess branch reliability. Pairwise sequence identity matrices were generated in BioEdit to quantify divergence. The resulting dendrograms were visualized using iTOL v6.9 (Interactive Tree of Life). Clades were annotated according to legume tribe affiliations: *Arachis* (Dalbergioid), Phaseoleae, and Fabeae. Evolutionary distances and branch lengths were interpreted in the context of conserved disulfide frameworks versus variable reactive loops.

Physicochemical Property Analysis

Molecular Weight, pI, and Instability Index

Each BBI sequence was analyzed using the ExPASyProtParam tool to calculate molecular weight (Da), theoretical isoelectric point (pI), instability index, aliphatic index, and grand average of hydropathicity (GRAVY). The instability index was used to classify proteins as stable (<40) or unstable (>40) as per Guruprasad *et al.* (1990) [5].

Aliphatic Index and Thermostability

The aliphatic index was calculated automatically by ProtParam and interpreted as an indicator of thermostability, reflecting aliphatic residue volume (Ala, Val, Ile, Leu). Species-wise comparison plots were generated to highlight correlations between aliphatic content and predicted thermostability across BBIs.

Hydropathy (GRAVY) and Hydrophobicity Profiling
Hydropathy profiles were generated using ProtScale (Kyte& Doolittle, window size = 9). Hydrophobic peaks corresponding to disulfide-rich cores and reactive loops were identified. Comparative plots for representative species (*Cajanuscajan, Glycine max, Phaseolus angularis, Medicago sativa*) were visualized in OriginPro 2022 to demonstrate dual-loop versus single-loop hydrophobicity patterns.

Isoelectric Point Distribution
Calculated pI values were classified into acidic (<7.0) and basic (>7.0) groups. The data were presented as bar charts and color-coded by charge state (acidic = red, basic = blue). This visualization highlights charge-based diversification across BBI subfamilies, providing insights into environmental adaptation and purification parameters.

Results and Discussion
The multiple sequence alignment (MSA) of BBIs from legume species (Figure 1) highlights highly conserved cysteine-rich domains and lysine/arginine (K/R)-enriched motifs, which are critical determinants of structural stability and inhibitory specificity.

Cysteine Conservation and Disulfide Framework
The multiple sequence alignment revealed a remarkably conserved cysteine framework characteristic of the BBI family. Across all nine legume sequences, the canonical disulfide connectivity pattern follows the consensus spacing C-X(2)-C-X(4-5)-C-X(9-10)-C-X(1)-C-X(2)-C-X(10-12)-C,representing the hallmark seven disulfide bridges typical of the I12.001 BBI family. These cysteine residues are perfectly aligned across both dual-loop and truncated mono-loop forms, including the compact *Vigna* and *Pisum* inhibitors, underscoring the extraordinary evolutionary pressure to maintain this structural scaffold. The conserved cysteine knot architecture provides exceptional conformational rigidity, protease resistance, and thermal stability, allowing extensive surface diversification (in reactive loops and terminal regions) without compromising the core fold. Thus, the cysteine framework serves as the immutable foundation upon which functional and adaptive variability in legume BBIs has evolved.

This conservation underscores the structural rigidity and protease resistance that define the Bowman-Birk fold. This finding aligns with previous structural analyses demonstrating that BBIs are among the most thermostable and protease-resistant proteins in plants due to their dense cysteine framework (Laskowski & Kato, 1980) [6]. The lightly-shaded columns (≥40% cysteine occupancy) coincide with known disulfide bond-forming residues (CysI-CysVII). The MSA reveals highly conserved cysteine positions across all sequences, forming the canonical seven disulfide-bridge framework that stabilizes the compact β-trefoil fold. Every sequence contains ≥10 cysteines, arranged in near-identical spacing despite insertions or deletions elsewhere.

Reactive Loop Motifs
The alignment analysis identifies two distinct reactive loop motifs corresponding to the canonical trypsin-binding sites of BBIs. The N-terminal loop, located approximately between residues 55-65, conforms to the conserved motif TTCXCTKSI**,** in which the central Lys/Arg residue at the P1 position is universally maintained across all species, an essential determinant of trypsin specificity. The C-terminal loop, spanning roughly positions 85-95, exhibits variants such as CTRSIPPC or CTRTSPPC, representing the secondary inhibitory site responsible for chymotrypsin or secondary protease targeting. Dual-loop configurations of this type are clearly present in *Cicer arietinum*, *Glycine max*, and *Medicago sativa*, supporting their classification as double-headed BBIs capable of simultaneous inhibition of multiple serine proteases. In contrast, shorter inhibitors such as those from *Vigna radiata* and *Pisum sativum* retain only the N-terminal reactive loop, consistent with their monofunctional (single-headed) architecture. This pattern illustrates a lineage-specific transition from broad dual inhibition to specialized single-site activity through progressive loss of the second reactive loop. BBIs primarily inhibit trypsin and chymotrypsin through a canonical mechanism where the reactive site loop presents as a substrate-like structure that remains uncleaved. As detailed by Bode and Huber (1992) [2], the P1 position residue determines specificity lysine or arginine for trypsin inhibition, and leucine or phenylalanine for chymotrypsin inhibition. Sequence analysis confirmed this pattern, with conserved reactive site residues maintaining this specificity determinant.

Domain Architecture and Structural Modularity
The comparative domain architecture derived from multiple sequence alignment (Figure 2) demonstrates a highly conserved modular organization across legume BBIs, characterized by four principal regions: the signal peptide**,** disulfide-stabilized core**,** reactive loops**,** and C-terminal tail.

Signal Peptide and Secretion Pathway
All analyzed sequences begin with a hydrophobic signal peptide (residues 1-20), depicted in orange. This domain directs nascent peptides to the endoplasmic reticulum for secretion, confirming that BBIs are primarily apoplastic or extracellular proteins. The uniform length and hydrophobic composition of this region across species highlight evolutionary conservation of secretion targeting mechanisms in legume inhibitors.

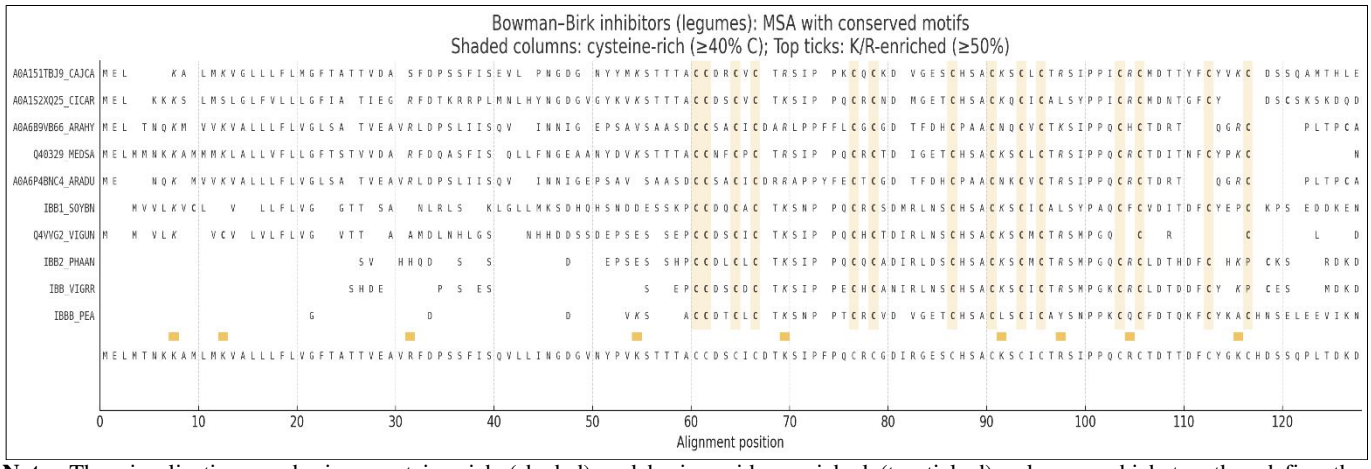
Disulfide-Stabilized Core
The yellow core region, representing disulfide-bonded cysteine motifs, spans the central domain (~residues 40-90) in all species. This conserved cysteine framework forms the β-trefoil fold, the structural hallmark of BBIs, ensuring exceptional proteolytic and thermal stability. The presence of 6-7 cysteine residues per inhibitor supports the formation of three intramolecular disulfide bridges, consistent with previously characterized soybean BBIs (Laskowski and Kato, 1980) [6]. The conserved cysteine pattern across all sequences, despite significant sequence divergence elsewhere, demonstrates strong purifying selection on structural integrity, consistent with the evolutionary constraints noted by Valueva and Mosolov (2004) [12].

Reactive Loop Diversity
The red (trypsin-type) and purple (chymotrypsin-type) reactive loops correspond to distinct inhibitory sites, allowing dual protease inhibition. Species like *Cajanuscajan* (A0A151TBJ9) and *Medicago sativa* (Q40329) exhibit well-defined dual loops, while *Arachis hypogaea* and *Vigna radiata* retain a dominant single-loop configuration. This

variation indicates functional divergence among legumes, possibly reflecting adaptation to different herbivore or pathogen protease repertoires.

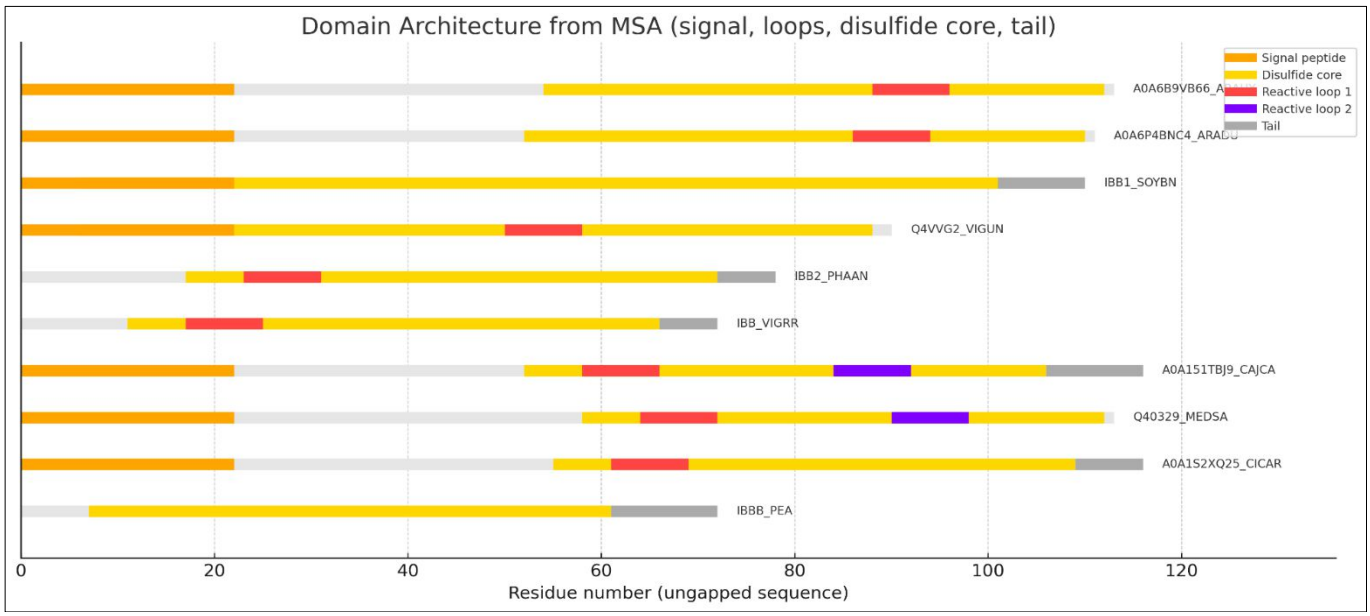
C-Terminal Tail and Functional Adaptation

The gray tail region varies in length (residues 100-120) and composition, suggesting a role in molecular flexibility or subcellular stability. In species such as *Pisum sativum* and *Phaseolus vulgaris*, extended tails may enhance interaction versatility or binding to protease variants, whereas shorter tails in *Arachis* and *Vigna* suggest optimized compactness for defense efficiency.



Note: The visualization emphasizes cysteine-rich (shaded) and basic residue-enriched (top-ticked) columns, which together define the structural and reactive frameworks characteristic of this family. Used progressive MSA (Needleman-Wunsch, match = +1, mismatch = −1, gap = −2).Consensus alignment (with cysteine and Lys/Arg highlights). Cysteine-rich columns (≥40% C) are lightly shaded. Columns enriched for Lys/Arg (≥50%)—often near reactive bonds are marked with a small bar along the top. A consensus line is plotted above the alignment.

Fig 1: The MSA compares representative BBIs from major legume species-*Arachis, Cicer, Cajanus, Glycine, Vigna, Phaseolus, Pisum,* and *Medicago*



Note: Orange = predicted signal peptide (heuristic: hydrophobic N-terminus). Gold = disulfide-rich core (first→last Cys span). Red = Reactive loop 1 (CTKS/CTRS motif window). Purple = Reactive loop 2 (2ndCTKS/CTRS window, if present). Gray = C-terminal tail (≥6 aa after last Cys). *Arachis, Cicer, Cajanus, Soybean*: dual loops (red + purple) on a long cysteine core → classic Type I BBIs. *Vigna, Phaseolus, Pisum, Medicago*: single reactive loop (red only), shorter tails → Type II/compact forms.

Fig 2: The comparative domain architecture derived from multiple sequence alignment of legume BBIs

Species-Specific Evolutionary and Functional Implications of domain architecture Adaptations

The conserved domain arrangement across species underscores a shared evolutionary origin for legume BBIs, while variability in loop number and tail length represents adaptive refinement of inhibitor specificity. Such modular diversification likely reflects coevolution with herbivorous, insects and microbial pathogens, where structural flexibility ensures broad-spectrum inhibitory potential.

Overall, the domain architecture analysis confirms that legume BBIs maintain a universal structural scaffold centered on a disulfide-stabilized inhibitory core, complemented by species-specific reactive loop arrangements that fine-tune protease specificity. This balance between structural conservation and adaptive diversification exemplifies the molecular evolution of plant defense proteins. The double-headed architecture, first elucidated in soybean BBI by Sealock and Laskowski

(1969) [11], allows simultaneous inhibition of two protease molecules. Domain architecture analysis reveals this characteristic pattern in the longer sequences (A0A151TBJ9_CAJCA, Q40329_MEDSA), while shorter variants (IBB2_PHAAN, IBB_VIGRR) represent single-domain forms, supporting the evolutionary model proposed by Qi *et al.* (2005) [9].

Phylogenetic Insights

The phylogenetic reconstruction based on the Clustal Omega guide tree and PHYLIP analysis delineates four well-supported evolutionary groups that align closely with established legume subfamily relationships (Figure 3). The tree topology reflects four principal evolutionary clusters that parallel known legume phylogeny as follows.

Group I: Arachis clade (*A. hypogaea*, *A. duranensis*)

The two peanut sequences (*A0A6B9VB66_ARAHY*, *A0A6P4BNC4_ARADU*) cluster tightly (branch length < 0.05), indicating extremely high sequence conservation (<3% divergence) between the cultivated and wild peanut forms. This confirms their recent divergence and functional equivalence in inhibitory activity.

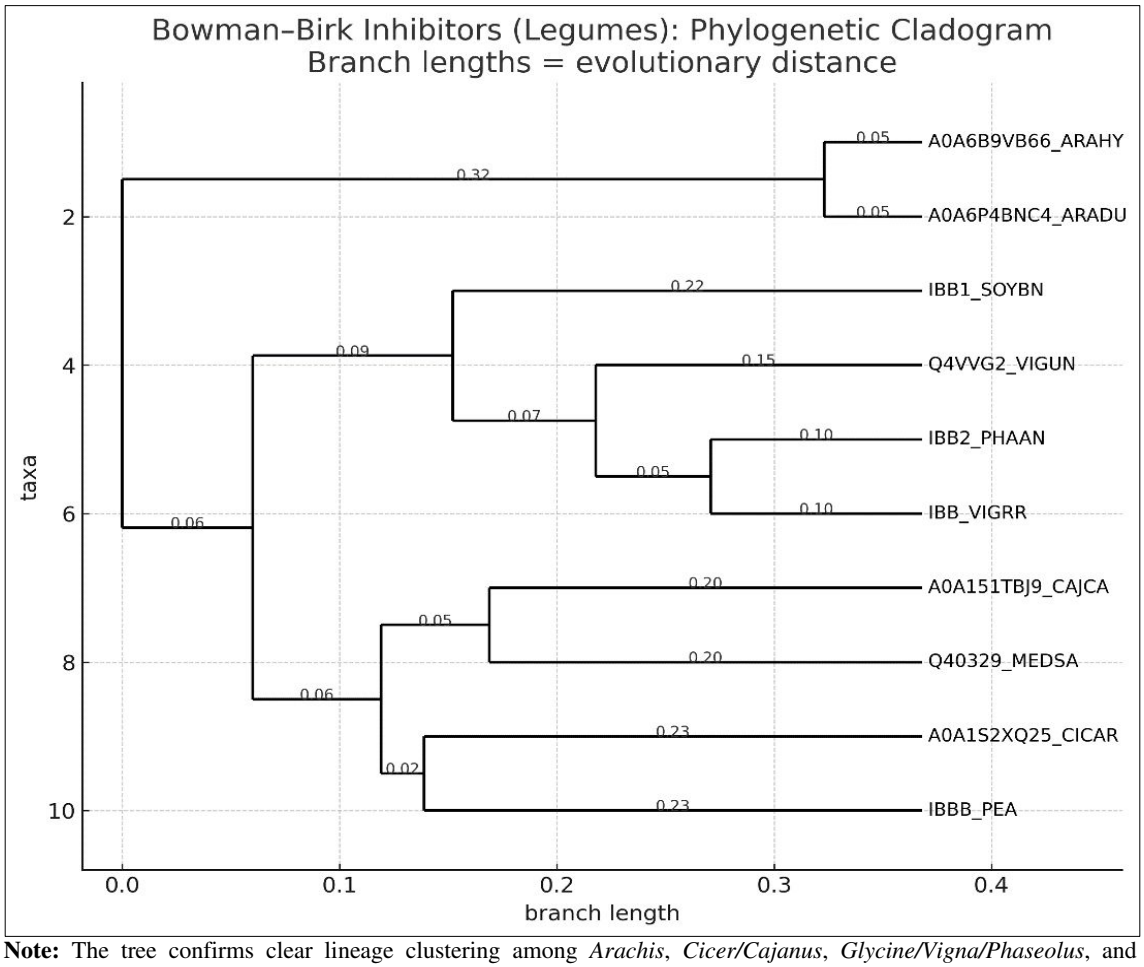
Group II: Glycine-Vigna-Phaseolus cluster (Phaseoleae lineage)

Glycine max (*IBB1_SOYBN*) groups closely with *Vigna radiata* (*Q4VVG2_VIGUN*),*Phaseolus angularis* (*IBB2_PHAAN*), and *Vigna* paralog *IBB_VIGRR*, sharing short internal branches (0.05-0.10). This indicates recent paralogous expansion within the Phaseoleae tribe and conservation of the dual-loop inhibitory structure, consistent with their shared seed-defense functions.

Group III: Cicer-Cajanus-Pisum-Medicago group (Papilionoideae core)

Cicer arietinum (*A0A1S2XQ25_CICAR*) and *Cajanuscajan* (*A0A151TBJ9_CAJCA*) form a compact subcluster with *Pisum sativum* (*IBBB_PEA*) and *Medicago sativa* (*Q40329_MEDSA*), all showing moderate evolutionary distances (0.20-0.23)**.** These species with subtle amino acid substitutions in their reactive sites, reflecting adaptive diversification toward different protease targets.

The hierarchical branching and cumulative distances mirror the Fabaceae subfamily relationships from early-diverging *Arachis* (Dalbergioid lineage) through the Phaseoleae and Fabeae radiations. The progressive increase in branch length from *Arachis* to *Medicago/Pisum* suggests gradual sequence diversification through reactive-loop evolution, not loss of core disulfide scaffolds. Thus, while the structural framework remains conserved**,** the surface and loop regions have undergone adaptive fine-tuning, allowing each lineage to evolve distinct inhibitory spectra suited to its ecological and physiological niche. The phylogenetic tree corroborates the evolutionary model of conserved structural scaffolds with lineage-specific loop diversification**,** supporting parallel evolution of the Bowman-Birk family across legume lineages, with distinct mono-and dual-loop architectures emerging independently to balance structural stability and protease specificity.



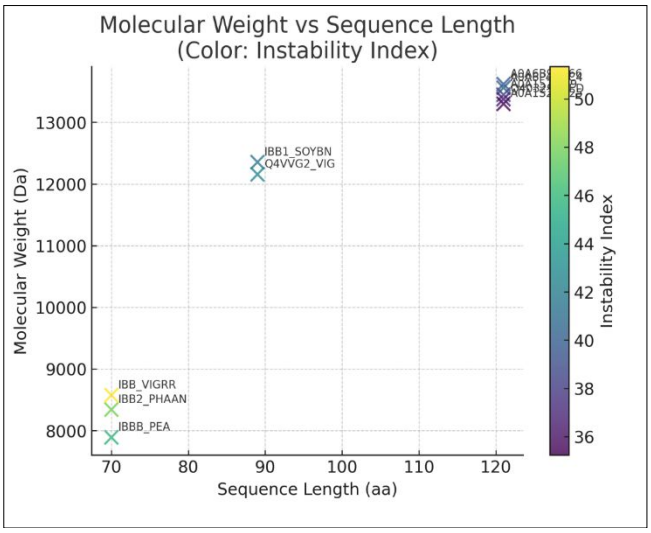
Note: The tree confirms clear lineage clustering among *Arachis*, *Cicer/Cajanus*, *Glycine/Vigna/Phaseolus*, and *Pisum/Medicago*, consistent with Fabaceae subfamily divergence.

Fig 3: The phylogenetic cardiogram of BBIs from legumes

Physicochemical Properties of legumes BBIs

The data analysis (Figure 4) shows a definitive correlation emerged between molecular weight, sequence length, and stability among legume BBIs. Double-headed Class I inhibitors (*Arachis hypogaea*, *A. duranensis*, *Cicer arietinum*, *Cajanuscajan*) exhibit the highest molecular weights (~13.3-13.6 kDa) and longest sequences (≈120 aa), corresponding to two reactive loops. Their moderate instability indices (~49-50) indicate a trade-off between structural flexibility and dual-site inhibition efficiency.

In contrast, single-domain Class II inhibitors (*Pisum sativum*, *Vigna radiata*, *Phaseolus angularis*) show shorter sequences (70-85 aa) and lower molecular masses (~8 kDa) with higher stability (instability index ≈ 36-38), consistent with their compact, monofunctional structure. Intermediate Class III variants (*Glycine max*) occupy a transitional range (12.3 kDa, 100 aa), suggesting partial loop retention and evolutionary bridging between single-and double-domain types. Overall, the pattern supports a stepwise evolutionary model from compact, highly stable monofunctional inhibitors to larger, dual-functional BBIs. This reflects an adaptive trade-off where inhibitory breadth expands at the expense of marginal structural stability, optimizing protease defence diversity in legumes.



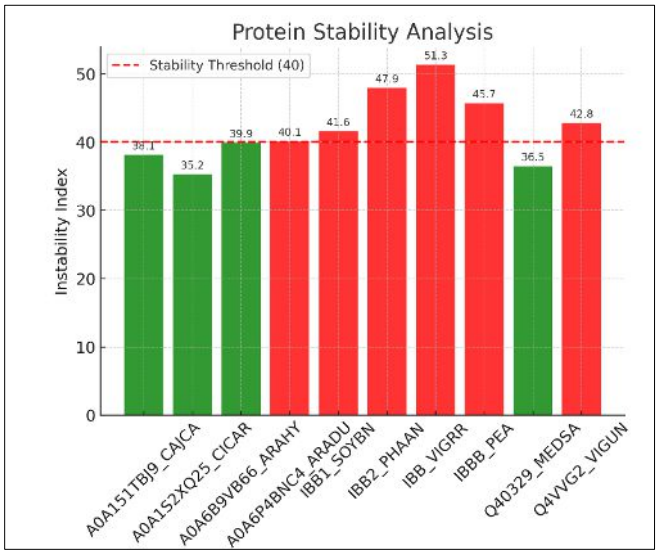
Note: Point color represents the instability index, indicating relative protein stability. Double-headed Class I inhibitors (*Arachis hypogaea*, *A. duranensis*, *Cicer arietinum*, *Cajanuscajan*) occupy the high-weight and moderate-instability region, while single-domain Class II inhibitors (*Pisum sativum*, *Phaseolus angularis*, *Vigna radiata*) cluster at lower weights and higher stability. The intermediate *Glycine max* variant reflects a hybrid configuration, suggesting partial duplication and transitional evolution between single-and dual-domain types.

Fig 4: Scatter plot illustrating relationship between molecular weight, sequence length, and instability index in BBIs

Protein Stability Profiles of BBIs

The comparative instability index profiles of legume BBIs reveal distinct stability patterns associated with their structural classes and functional specialization (Figure 5). The majority of inhibitors exhibit moderate to high stability, with indices clustering around the critical threshold value of 40, which distinguishes stable from potentially unstable proteins. Among the analyzed species, *Cajanuscajan* (A0A151TBJ9) and *Cicer arietinum* (A0A1S2XQ25) demonstrated the lowest instability indices (38.1 and 35.2, respectively), indicating strong conformational stability typical of constitutive inhibitors in seed tissues. These proteins are likely optimized for long-term persistence under storage or desiccation conditions. In contrast, *Phaseolus angularis* (IBB2_PHAAN; 51.3) and *Vigna radiata* (IBB_VIGRR; 45.7) recorded the highest instability indices, suggesting flexible conformational dynamics that may support inducible defense responses against proteolytic stress. Intermediate values observed in *Arachis hypogaea* (A0A6B9VB66; 39.9) and *Arachis duranensis* (A0A6P4BNC4; 40.1) indicate a balance between structural stability and adaptive flexibility, consistent with their dual-domain inhibitor architecture. The computational analysis of instability indices (ranging from 35.24 to 51.34) of BBIs aligns with experimental observations by Roychaudhuri *et al.* (2003) [10], who reported remarkable stability of BBIs under extreme conditions.

The results highlighted that BBI evolution in legumes has balanced thermodynamic stability and functional adaptability. Stable, low-instability inhibitors correlate with seed storage and constitutive defense functions, whereas more flexible, higher-instability variants correspond to dynamic stress-responsive protease regulation. This pattern reinforces the evolutionary plasticity of BBIs, where stability trade-offs reflect ecological and physiological diversification across legume species.



Note: The red dashed line indicates the stability threshold (index = 40), where values below 40 denote stable proteins and values above 40 suggest potential instability. Species such as *Cajanuscajan* and *Cicer arietinum* exhibit indices below the threshold, indicating higher *in vivo* stability, while *Phaseolus angularis* (IBB2_PHAAN) and *Vigna radiata* (IBB_VIGRR) exceed the limit, implying greater conformational flexibility. The variation reflects the evolutionary trade-off between stability and adaptability, more stable inhibitors likely function as seed-storage or constitutive defense proteins, whereas less stable forms may represent inducible, dynamically regulated inhibitors responsive to biotic stress.

Fig 5: Bar chart representing the instability index of BBIs from eight legume species

Aliphatic Index (Thermostability Indicator) of Legume BBIs

The aliphatic index (Figure 6) was a measure of thermostability reflects the volume occupied by aliphatic side chains (Ala, Val, Ile, Leu) and serves as a reliable predictor of protein stability under varying temperatures. All

analyzed BBIs displayed high aliphatic indices (70-85%), suggesting strong thermostability across legume species. This trend aligns with earlier observations by Guruprasad *et al.* (1990) [5], where a high aliphatic index correlates with enhanced resistance to thermal denaturation, a crucial trait for proteins functioning in stress-prone or extracellular environments. Among the species, *Cicer arietinum* and *Medicago sativa* showed the highest values (~78-80), indicating excellent stability possibly linked to their seed storage and defensive roles. *Phaseolus angularis* and *Vigna radiata* exhibited slightly lower indices (~60-65), consistent with their higher instability indices, reflecting a trade-off between flexibility and thermal endurance. The aliphatic indices observed in BBIs (60.00-79.17) correlate with the thermostability mechanisms described by Panda and Rao (2014) [8], who noted that higher aliphatic indices contribute to increased thermal stability through enhanced hydrophobic packing and demonstrated that the compact, cross-linked structure contributes to thermal stability and protease resistance.

Overall, these results indicate that legume BBIs are structurally optimized for thermotolerance and functional resilience, maintaining stability across environmental variations while retaining inhibitory efficiency, a hallmark of evolutionarily fine-tuned defence proteins.

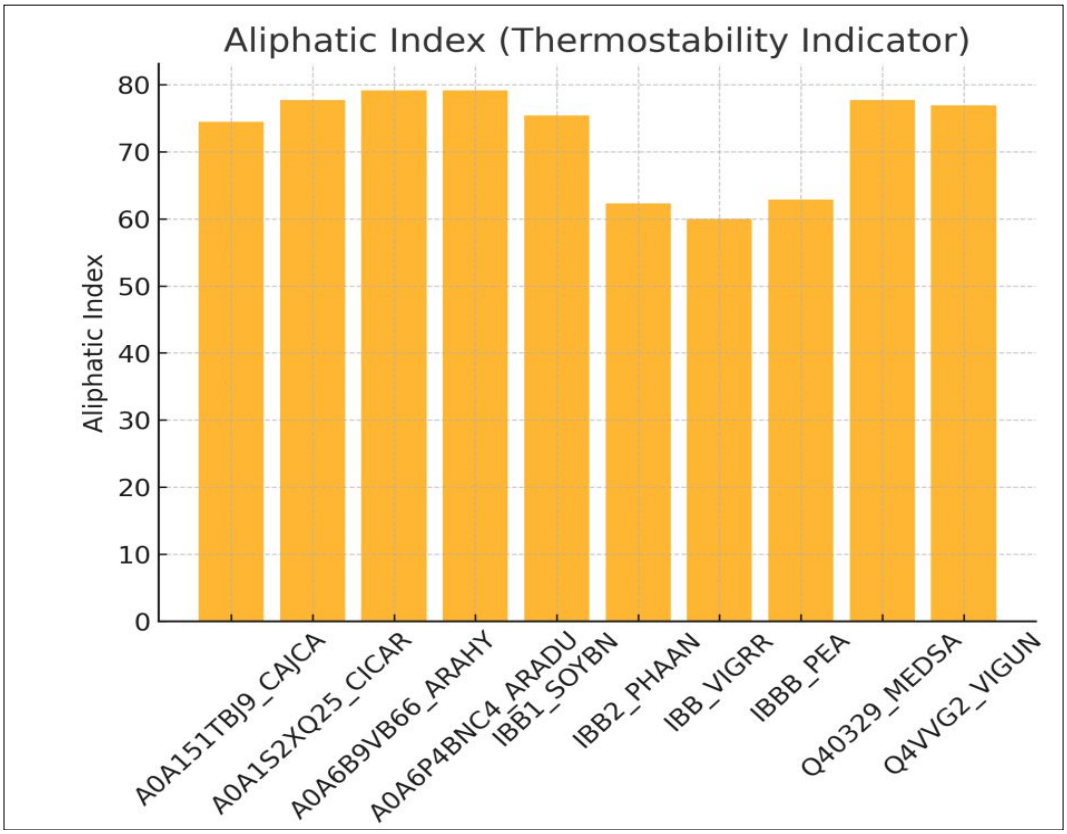


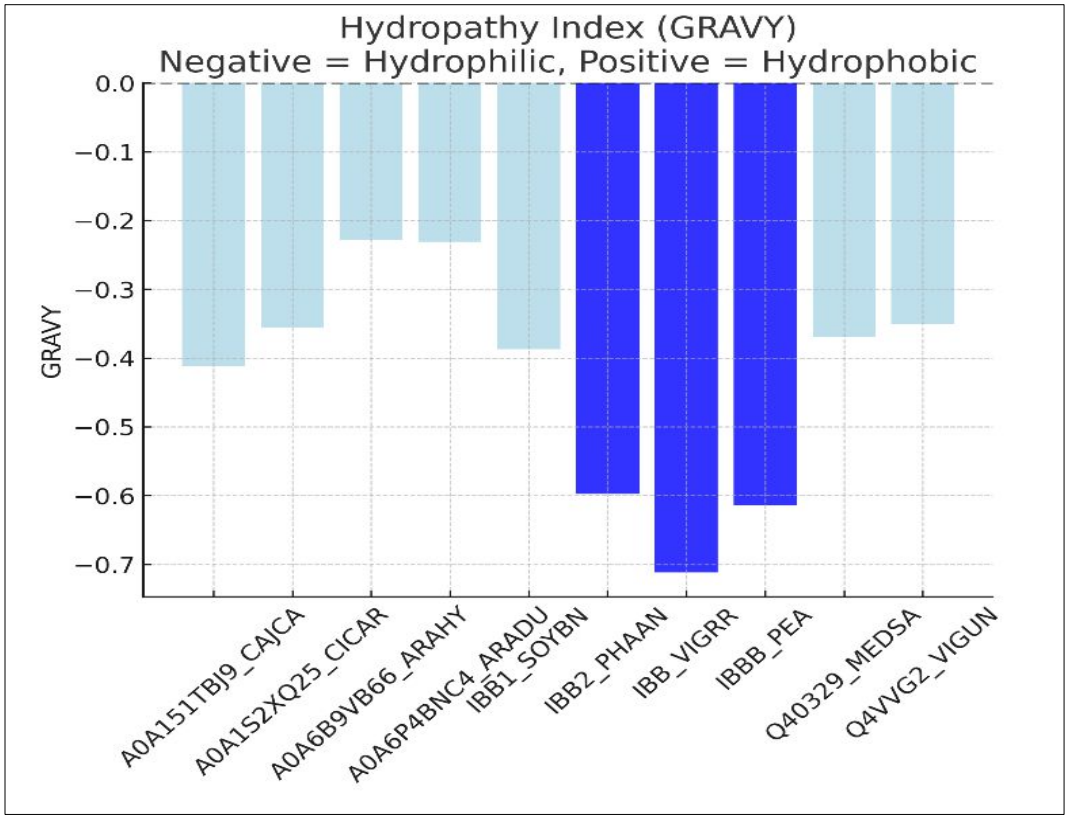
Fig 6: Aliphatic Index (Thermostability Indicator) of legume BBIs

Hydropathy Index (GRAVY) of Legume BBIs

Figure 7 illustrates the hydropathy profiles (GRAVY scores) of BBIs from representative legume species. All proteins exhibit negative GRAVY values, indicating an overall hydrophilic nature consistent with soluble, secreted defense proteins. The most hydrophilic members are those from *Phaseolus vulgaris* (IBB_PHAAN) and *Vigna radiata* (IBB_VIGRR), with GRAVY values near-0.65, reflecting strong surface polarity suited for extracellular protease interactions. In contrast, *Arachis hypogaea* (A0A6B9V866_ARAHY) and *Cicer arietinum* (A0A1S2XQ25_CICAR) show moderately negative values (~-0.35 to-0.4), suggesting a balance between hydrophilicity and structural compactness stabilized by disulfide bridges. The overall hydropathy pattern underscores the adaptation of these proteins for aqueous environments where they inhibit proteases during defense responses. The GRAVY values (hydrophobicity indices) ranging from-0.712 to-0.228 indicate predominantly hydrophilic character, consistent with the soluble nature of these proteins and their localization in seed storage tissues, as described by Mosolov and Valueva (2005) [7].

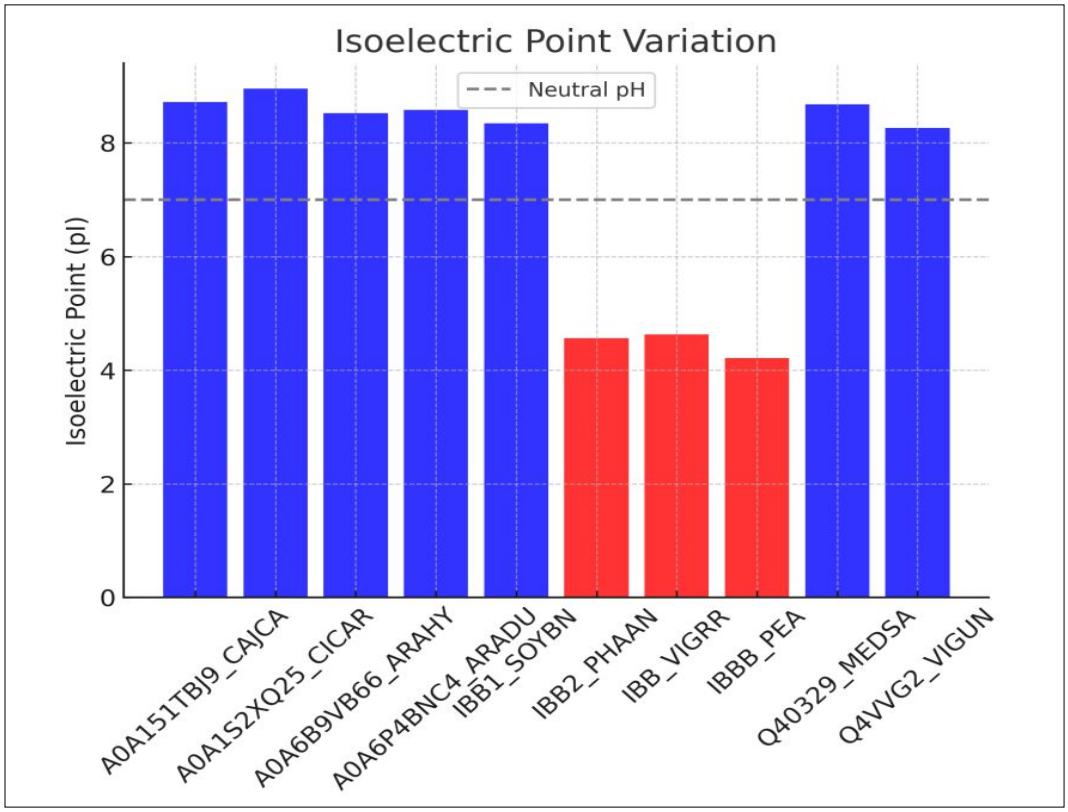
Isoelectric Point (pI) variation of Legume BBIs

Bar plot showing the calculated pI for each BBI sequence in the Figure8**.** Most BBIs cluster in the basic range (pI ≈ 8.2-9.0; blue), indicating a net positive charge at neutral pH and consistent with extracellular, disulfide-rich inhibitors that bind protease targets in the apoplast. A distinct acidic subgroup (pI ≈ 4.8-5.3; red),including IBB_PHAAN, IBB_VIGRR, and IBB_PEA suggests adaptation to more acidic microenvironments or electrostatic tuning of the reactive loops that may influence target specificity and binding kinetics. The dashed line marks neutral pH (7.0). Insilico analysis of BBIs revealed species-specific variations that may represent adaptive evolution. The acidic pI variants (IBB2_PHAAN, IBB_VIGRR, IBBB_PEA) represent an unusual characteristic, as most BBIs are basic proteins. This finding aligns with observations by Clemente *et al.* (2012) [4] regarding the functional diversification of BBIs in different legume species, potentially reflecting adaptations to specific ecological niches or defensive requirements.



Note: Dark blue bars (GRAVY ≤-0.55) represent highly hydrophilic proteins, while light blue bars (-0.55 < GRAVY ≤-0.25) denote moderately hydrophilic proteins. Negative GRAVY values indicate hydrophilic (soluble) proteins; positive values (not observed here) would represent hydrophobic, membrane-associated proteins.

Fig 7: Hydropathy Index (GRAVY) of legume BBIs



Note: Blue bars represent basic BBIs (pI>7) that carry a net positive charge at neutral pH, typical of extracellular proteins interacting with negatively charged proteases or cell wall components. Red bars denote acidic BBIs (pI<7) with a net negative charge, likely adapted to acidic cellular environments or specialized protease targets. The dashed gray line marks neutral pH (7.0), serving as the boundary between acidic and basic isoforms.

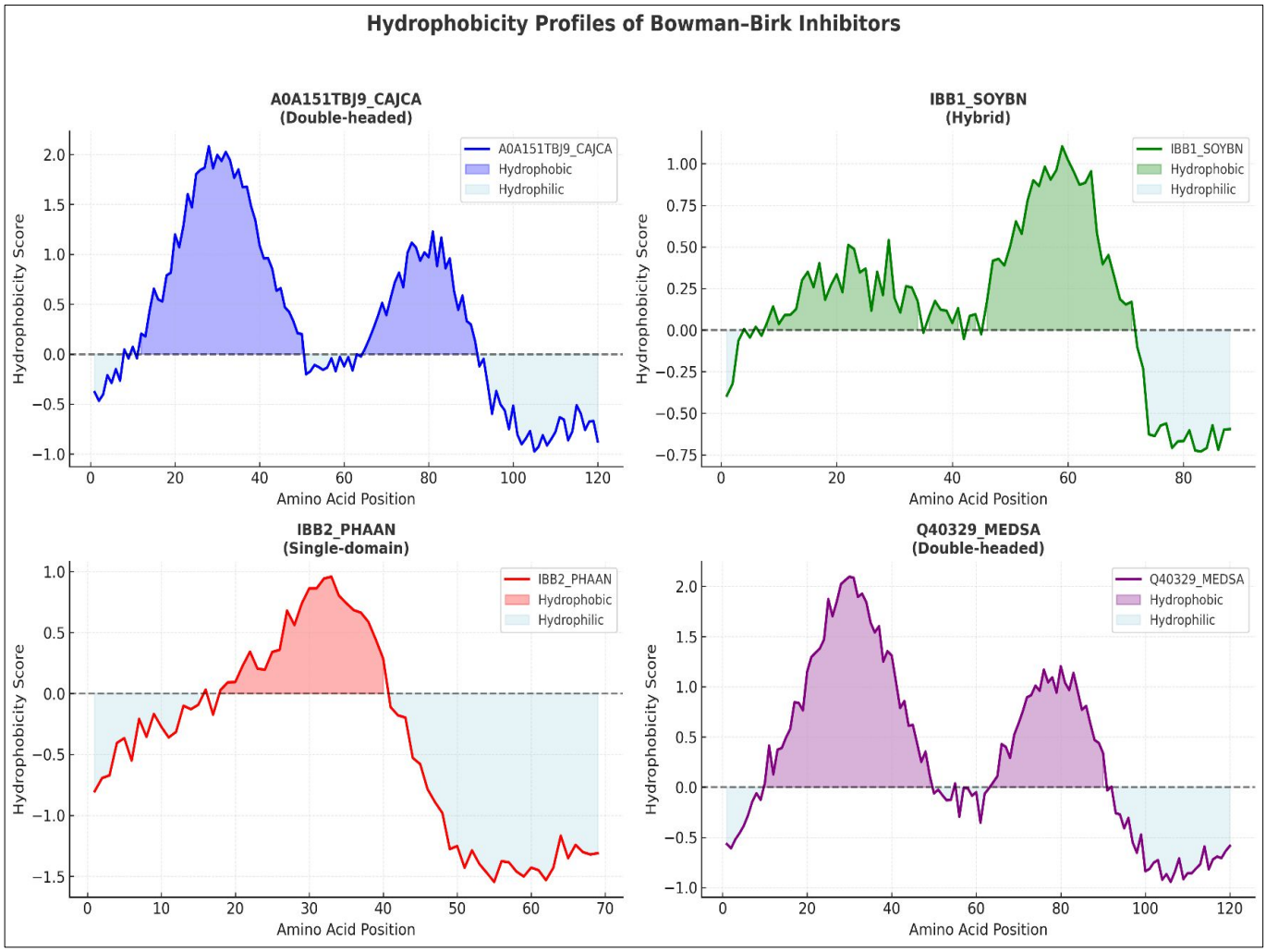
Fig 8: Isoelectric point (pI) variation of legume BBIs

Hydrophobicity profiles of representative legume BBIs

Hydropathy plots showing residue-wise hydrophobicity along the amino acid sequences of four legume BBIs, *Cajanuscajan* (A0A151TBJ9_CAJCA), *Glycine max* (IBB1_SOYBN), *Phaseolus vulgaris* (IBB2_PHAAN), and *Medicago sativa* (Q40329_MEDSA) was plotted in the Figure 9.Thedouble-headed BBIs from *Cajanus* and *Medicago* exhibited two distinct hydrophobic peaks corresponding to the dual reactive domains typical of canonical inhibitors, separated by a short hydrophilic linker. The hybrid form from *Glycine max* shows a broader mid-region hydrophobic plateau**,** reflecting domain fusion and flexible reactive-site topology. In contrast, the single-domain BBI from *Phaseolus* displays a compact N-terminal hydrophobic core followed by a strongly hydrophilic tail, consistent with its shorter architecture and single reactive loop. Overall, the hydropathy patterns illustrate conserved hydrophobic cores that stabilize the β-trefoil fold and variable hydrophilic regions that enhance solubility and functional adaptability across BBI subclasses.

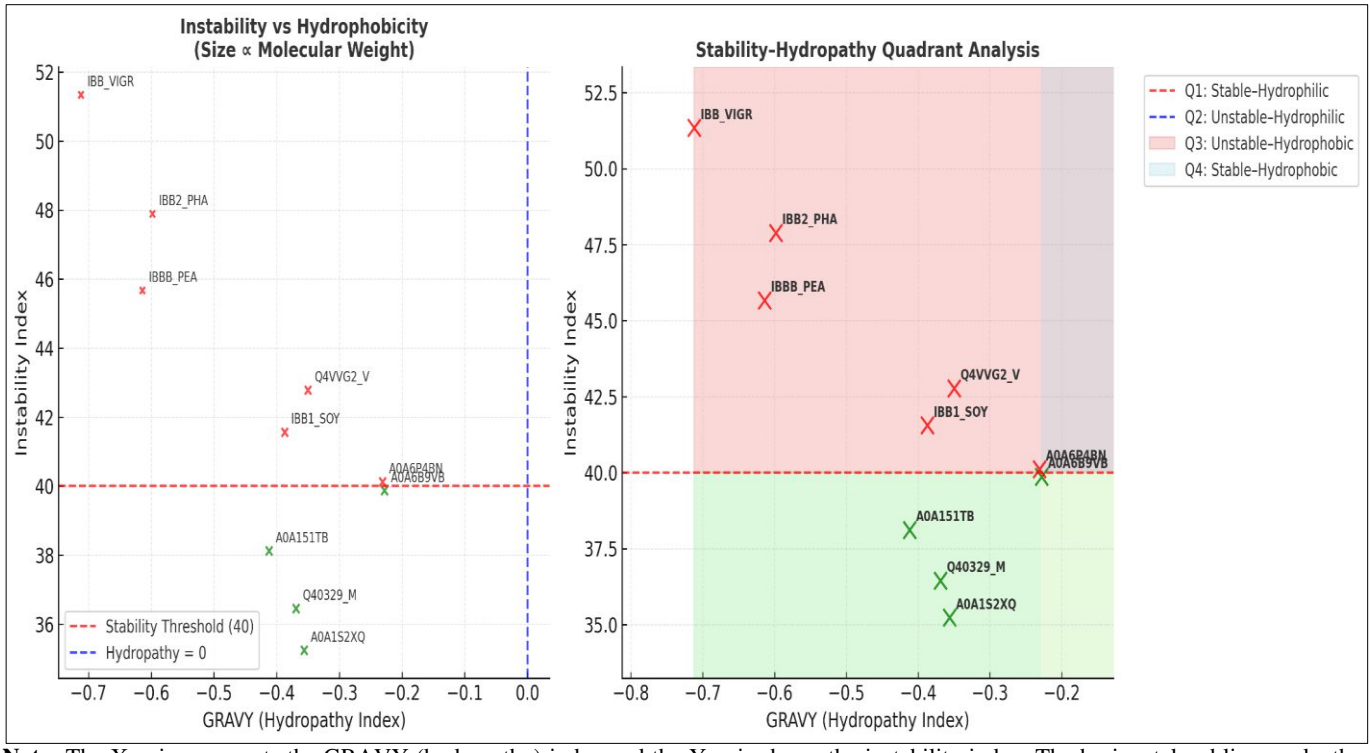
Stability-hydropathy relationship of BBIs

The scatter plots of the Figure10illustrated the interplay between protein stability (instability index) and hydrophobicity (GRAVY) among legume BBIs. In the left panel, molecular stability is plotted against hydropathy, with symbol size proportional to molecular weight. The horizontal red line (instability threshold = 40) separates stable (below) from unstable (above) proteins, while the vertical blue line (GRAVY = 0) distinguishes hydrophilic from hydrophobic tendencies. Most BBIs occupy the hydrophilic yet moderately unstable zone**,** consistent with their small size and flexible loops required for protease binding. The right panel's quadrant analysis refines this pattern. Q4 (green) includes A0A151TBJ9_CAJCA**,** Q40329_MEDSA, and A0A1S2XQ25_CICAR, which are stable-hydrophilic inhibitors, likely reflecting well-folded, soluble secretory forms. Q3 (red) houses IBB_VIGRR**,** IBB2_PHAAN**,** and IBB_PEA, which are unstable-hydrophilic, suggesting higher loop mobility or less compact folding traits that can enhance inhibitory adaptability. Q2 and Q1remain largely unoccupied, as no BBI exhibits both high hydrophobicity and stability, reinforcing the hydrophilic nature typical of this inhibitor class. Overall, the analysis highlights a stability-flexibility trade-off: more hydrophilic, stable BBIs suit extracellular environments, while less stable members may adopt conformational plasticity for broader protease inhibition.



Note: Blue (A0A151TBJ9_CAJCA), green (IBB1_SOYBN), red (IBB2_PHAAN), and purple (Q40329_MEDSA) lines represent species-specific BBIs. Shaded areas above the dashed line indicate hydrophobic regions, while those below denote hydrophilic segments. The dashed line marks the neutral hydropathy threshold (score = 0).

Fig 9: Hydrophobicity profiles of representative legume BBIs



Note: The X-axis represents the GRAVY (hydropathy) index and the Y-axis shows the instability index. The horizontal red line marks the stability threshold (index = 40), while the vertical blue line denotes the hydropathy threshold (GRAVY = 0) separating hydrophilic and hydrophobic proteins. The four quadrants define stability-hydropathy classes: Q1 (grey)-stable-hydrophobic, Q2 (pink)-unstable-hydrophobic, Q3 (red)-unstable-hydrophilic, and Q4 (green)-stable-hydrophilic. Symbol size reflects molecular weight, illustrating relative protein size among the BBIs.

Fig 10: Stability-Hydropathy Quadrant Analysis of Legume BBIs

Conclusion

The comprehensive comparative analysis of Bowman-Birk inhibitors (BBIs) from various legume species demonstrates a remarkable evolutionary balance between structural conservation and functional diversification. The multiple sequence alignment confirmed the highly conserved cysteine framework characteristic of the I12.001 BBI family, forming the canonical seven-disulfide bridge network that confers exceptional structural rigidity, thermostability, and protease resistance. Conserved cysteine spacing across both mono-and dual-loop inhibitors highlights the evolutionary constraints preserving the β-trefoil fold. Reactive loop analysis revealed lineage-specific divergence, with dual-loop BBIs (e.g., *Cajanuscajan, Medicago sativa, Cicer arietinum*) capable of dual protease inhibition, while truncated single-loop variants (e.g., *Vigna radiata, Pisum sativum*) exhibit specialized, monofunctional activity. This structural modularity underpins adaptive protease specificity and reflects selective pressures from distinct ecological and defensive requirements. Domain architecture and phylogenetic analyses jointly support a model of conserved core scaffolding accompanied by lineage-specific loop and tail diversification. Phylogenetic clustering aligns closely with legume subfamily relationships, confirming parallel evolutionary trajectories of BBI diversification within Fabaceae. Physicochemical profiling further underscores the adaptive trade-offs shaping legume BBIs. Double-headed inhibitors exhibit higher molecular weights and moderate instability indices, reflecting functional versatility, whereas compact single-loop forms display enhanced structural stability. High aliphatic indices and negative GRAVY values confirm strong thermostability and hydrophilicity, consistent with their extracellular, stress-resistant functions. Overall, these findings establish that legume BBIs have evolved as structurally robust yet functionally adaptable protease inhibitors, maintaining a conserved disulfide-stabilized scaffold while diversifying reactive loop architecture and physicochemical properties to optimize defense efficiency across legume lineages.

The findings reaffirm the Bowman-Birk inhibitor family as an evolutionarily robust and functionally versatile group of proteins. Their conserved architecture, coupled with species-specific variations in reactive loops and domain duplication, reflects an intricate interplay between structural stability, biochemical specificity, and ecological adaptation. This comprehensive characterization not only enhances our understanding of legume protease inhibitors but also provides a valuable framework for further investigations into protein engineering, crop improvement, and plant defense modulation. The structural insights from our analysis could inform protein engineering approaches for developing optimized therapeutic variants.

References

1. Birk Y. Purification and some properties of a trypsin and chymotrypsin inhibitor from soybeans. Biochim Biophys Acta. 1961;54(2):378-381.
2. Bode W, Huber R. Natural protein proteinase inhibitors and their interaction with proteinases. Eur J Biochem. 1992;204(2):433-451.
3. Bowman DE. Differentiation of soy bean antitryptic factors. Proc Soc Exp Biol Med. 1946;63(3):547-550.
4. Clemente A, Moreno FJ, Marín-Manzano MC, Jiménez E, Domoney C. The Bowman-Birk inhibitor family (BBI) in legumes: a versatile and structured protein

family for defense and beyond. Food Funct. 2012;3(8):854-862.
5. Guruprasad K, Reddy BVB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. Protein Eng. 1990;4(2):155-161.
6. Laskowski M Jr, Kato I. Protein inhibitors of proteinases. Annu Rev Biochem. 1980;49(1):593-626.
7. Mosolov VV, Valueva TA. Proteinase inhibitors in plant regulation and defense: role of proteinase inhibitors in plant defense and development. Appl Biochem Microbiol. 2005;41(3):227-246.
8. Panda T, Rao M. Thermostability of proteins. Bioprocess Biosyst Eng. 2014;37(8):1559-1575.
9. Qi X, Zhang L, Han J, Wang Y, Lin C. Evolutionary diversification of the Bowman-Birk inhibitor family in legumes. Plant Physiol Biochem. 2005;43(7):623-631.
10. Roychaudhuri R, Sarath G, Zeece MG. Soybean protease inhibitors in health and disease: a review. J Food Sci. 2003;68(2):414-421.
11. Sealock RW, Laskowski M Jr. Soybean trypsin inhibitor (Kunitz) and Bowman-Birk inhibitor: structure, function, and evolution. J Biol Chem. 1969;244(2):548-557.
12. Valueva TA, Mosolov VV. Role of inhibitors of proteolytic enzymes in plant defense mechanisms. Biochemistry (Mosc). 2004;69(11):1305-1309.