J Mol Evol (1996) 42:560–569

JOURNAL OF MOLECULAR EVOLUTION

Analysis of the Amino Acid Sequences of Plant Bowman-Birk Inhibitors

Balaji Prakash,[1] S. Selvaraj,[1,*] M.R.N. Murthy,[1] Y.N. Sreerama,[2] D. Rajagopal Rao,[2] Lalitha R. Gowda[2]

[1] Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India
[2] Department of Biochemistry and Nutrition, Central Food Technological Research Institute, Mysore 570002, India

Received: 13 March 1995 / Accepted: 27 November 1995

Abstract. Plant seeds contain a large number of protease inhibitors of animal, fungal, and bacterial origin. One of the well-studied families of these inhibitors is the Bowman-Birk family(BBI). The BBIs from dicotyledonous seeds are 8K, double-headed proteins. In contrast, the 8K inhibitors from monocotyledonous seeds are single headed. Monocots also have a 16K, double-headed inhibitor. We have determined the primary structure of a Bowman-Birk inhibitor from a dicot, horsegram, by sequential edman analysis of the intact protein and peptides derived from enzymatic and chemical cleavage. The 76-residue-long inhibitor is very similar to that of *Macrotyloma axillare*. An analysis of this inhibitor along with 26 other Bowman-Birk inhibitor domains (MW 8K) available in the SWISSPROT databank revealed that the proteins from monocots and dicots belong to related but distinct families. Inhibitors from monocots show larger variation in sequence. Sequence comparison shows that a crucial disulphide which connects the amino and carboxy termini of the active site loop is lost in monocots. The loss of a reactive site in monocots seems to be correlated to this. However, it appears that this disulphide is not absolutely essential for retention of inhibitory function. Our analysis suggests that gene duplication leading to a 16K inhibitor in monocots has occurred, probably after the divergence of monocots and dicots, and also after the loss of second reactive site in monocots.

Key words: Bowman-Birk — Inhibitors — Sequence analysis — Monocotyledon — Dicotyledon — Gene duplication — *Dolichos biflorus*

Introduction

Most plant seeds contain large amounts of inhibitors which appear to be mainly involved in defense against various parasites by inhibiting their digestive enzymes (Ryan 1981). One of the well-studied classes of serine protease inhibitors is the Bowman-Birk family of inhibitors (BBI), which are rich in cysteines. Many BBI sequences have been determined from both monocotyledonous and dicotyledonous seeds. BBIs from dicots usually have a molecular weight of 8K and are double headed (two reactive sites in a single inhibitor molecule) (Birk 1985), while those from monocots can be divided into two classes, one of size ~ 8K with one reactive site and the other of size ~ 16K with two reactive sites (Tashiro et al. 1987, 1990; Odani et al. 1986; Nagasue et al. 1988). The 16K proteins have two 8K BBI domains. In all these inhibitors, the residue toward the amino-terminal side of the scissile bond (P1) determines the specificity of inhibition. In this paper we present the primary structure of the horsegram (*Dolichos biflorus*) Bowman-Birk inhibitor and an analysis of amino acid sequences of 26 other BBI domains with a view to understand the role of cysteine residues, the variability of the active site sequence, and their implications in the molecular evolution of these inhibitors.

*S. Selvaraj is on leave from Department of Physics, Bharathidasan University, Tiruchirapalli 620 024, Tamilnadu, India
Correspondence to: M.R.N. Murthy

Materials and Methods

Purification of Horsegram Inhibitor (HGI). Horsegram seeds were obtained from the local market. These seeds contain several isoinhibitors belonging to the Bowman-Birk family. These inhibitors were purified from buffer extracts of defatted seed flour by fractional precipitation with ammonium sulphate, anion exchange chromatography, and gel filtration using Sephadex G-50 as reported earlier (Rama Sarma and Rajagopal Rao 1991). The four isoinhibitors present were resolved on a DEAE-Sephacel column using an ammonium bicarbonate gradient (Joubert et al. 1979). The predominant isoinhibitor HGI-III was used for protein sequencing.

Protein Sequencing. The inhibitor was reduced and either S-pyridylethylated (PE) by the method of Hermodson et al. (1973) or carboxymethylated by the method of Crestfield et al. (1963). The excess reagents were removed by reverse-phase HPLC on an Aquapore C-8,300Å (Pierce, U.S.A.) column using a linear gradient (0–100%) of CH_3CN in distilled water containing 0.1% trifluoroacetic acid (TFA). The PE-HGI-III (5 mg) was partially digested with lysyl endopeptidase at an enzyme/substrate ratio of 1/100 (w/w) in 0.1 M Tris-HCl (pH 8.5) at 37°C for 4 h. The CM-HGI-III (5 mg) was partially digested with TCPK-trypsin (1:100) in 0.1 M ammonium bicarbonate at 37°C for 10 min. The digests were lyophilized, dissolved in 0.1% TFA, and fractionated by reverse-phase HPLC on a Shimpak C-18 column (Shimadzu, Japan) using an appropriate 0.1% TFA/CH_3CN gradient (Mahoney et al. 1980). Most of the peptides were further purified by ion pair chromatography using the Shimpak C-18 or the Aquapore C-8 column (Yang et al. 1981). The CM-HGI-III was cleaved chemically using cyanogen bromide (Waxadal et al. 1968) and the peptides were purified as described above.

End-group analysis of the CM-HGI-III was performed using the 4-N,N'-dimethylaminoazobenzene 4'-isothiocyanate/phenylisothiocyanate double coupling method (Chang 1979). Carboxy-terminal analysis of the CM-HGI-III was performed by digestion with carboxypeptidase A as reported by Ambler (1972). The released amino acids were estimated after derivatization with phenylisothiocynate using the Waters Pico Tag amino acid analysis system (Bidlingmeyer et al. 1984). The pure peptides were sequenced using a Shimadzu PSQ-1 gas-phase sequenator.

Sequences. BBI sequences were obtained from the SWISSPROT databank, release June 1994, available at the Bioinformatics Center of the Indian Institute of Science, Bangalore, India. There were 25 sequences in the data bank which belonged to the BBI family. The sequence from maize was a wound-induced inhibitor of approximately 120 residues and was not included in the analysis. Thus, 24 sequences were chosen for analysis. The length of the sequences varied from 53 to 133 amino acids. The sequences from barley (Nagasue et al. 1988), wheat (Odani et al. 1986), and rice bran (Tashiro et al. 1987) are of lengths 124, 126, and 133 residues, respectively. BBI domains are usually 63–66 residues long. The sequences of both amino- and carboxy-terminal halves of rice bran and barley inhibitors are homologous to the 8K BBIs of monocots. Hence, these halves were treated as independent sequences in the multiple sequence alignment. Only partial sequence for the wheat inhibitor was available. The amino terminal segment of this sequence was treated as an independent 8K domain. Thus, a total of 27 domain sequences including that of HGI were analyzed (Table 1).

Sequence Comparisons. The pairwise sequence comparisons were carried out using the IALIGN program in the PRONUC suite of programs (Desai and Bourne 1986). IALIGN uses the Needleman and Wunsch algorithm (Needleman and Wunsch 1970) for deriving the optimal alignment of sequences. The match scores were based on Dayhoff's mutation data matrix (Schwartz and Dayhoff 1978). The matrix elements were made positive by addition of six. A penalty of six was used to minimize the gaps introduced in the alignment, irrespective of the length of the gaps. The significance of the alignments were obtained by randomization of the original sequences. One hundred such comparisons were used to evaluate the mean and standard deviation of the scores for unrelated sequences. The significance of the score for alignment of the native sequences was expressed in units of standard deviation above the mean score for unrelated sequences.

Table 1. List of sequences used

WHEAT1	Partially sequenced 16K BBI from wheat (*Triticum aestivum*)
WHEAT2	8K BBI from wheat (*Triticum aestivum*)
HORVUN	N-Terminal half of the 16K BBI from barley (*Hordeum vulgare*)
HORVUC	C-Terminal half of the 16K BBI from barley (*Hordeum vulgare*)
ORYSAN	N-Terminal half of the 16K BBI from rice (*Oryza sativa*)
ORYSAC	C-Terminal half of the 16K BBI from rice (*Oryza sativa*)
SETIT2	8K BBI from foxtail millet (FMTI-II) (*Setaria italica*)
SETIT3	8K BBI from foxtail millet (FMTI-III) (*Setaria italica*)
COILA1	8K BBI from job's tears (TI-1) (*Coix lachryma jobi*)
COILA2	8K BBI from job's tears (TI-2) (*Coix lachryma jobi*)
ARAHY1	8K BBI from peanut (A-II) (*Arachis hypogaea*)
ARAHY2	8K BBI from peanut (B-II) (*Arachis hypogaea*)
PHAAN	8K BBI from adzuki bean (*Phaseolus angularis; Vigna angularis*)
PHAAN2	8K BBI from adzuki bean (I-A) (*Phaseolus angularis; Vigna angularis*)
PHALU	8K BBI from lima bean (*Phaseolus lunatus*)
DOLAX3	8K BBI from *Macrotyloma axillaris* (DE-3)
DOLAX4	8K BBI from *Macrotyloma axillaris* (DE-4)
LONCA4	8K BBI from apple leaf seed (*Lonchocarpus capassa*)
PHAVU	8K BBI from kidney bean (*Phaseolus vulgaris*)
SOYBN	8K BBI from soybean (BBI) (*Glycine max*)
SOYBN2	8K BBI from soybean (C-II) (*Glycine max*)
SOYBN3	8K BBI from soybean (D-II) (*Glycine max*)
PHAAU	8K BBI from mung bean (*Phaseoulus aureus; Vigna radiata*)
VICAN	8K BBI from common vetch (VAI) (*Vicia angustifolia*)
VICFA	8K BBI from broad bean (FBI) (*Vicia faba*)
MEDSA	Wound induced 8K BBI from Alfalfa

Multiple Alignment and Construction of Phylogenetic Tree. The program MULTALIN (Corpet 1988) was used to align all the BBI sequences. This program has been found to produce near-optimal solutions for multiple sequence alignments. In this method, all possible pairs of sequences are aligned initially using Lipman and Pearson (1985) FASTP algorithm. The similarity scores obtained were then used to determine the hierarchical order of clustering sequences. The alignment of sequences that had the highest score was initially accepted and the aligned pair was treated as a single sequence. Each further step combined either two sequences or clusters or a sequence and a cluster. The similarity measure was reevaluated after each combination. This procedure was continued until all the sequences were merged. Aligned sequences were analyzed with the PHYLIP package of programs (Felsenstein 1988, 1989). The program SEQBOOT was used to produce 100 bootstrap replacement sequence alignments from the aligned sequences (Felsenstein 1985). These sequences were analyzed by the maximum parsimony method using the program PROTPARS. Trees produced by PROTPARS were analyzed by the program CONSENSE

to deduce a consensus tree. The unrooted phylogenetic tree was drawn using the program DRAWTREE. The number at the branches indicates the number of times that particular branch appeared in all the trees that were used to determine the consensus tree. These numbers varied from 30 to 100%, with a mean value of 80.

Construction of Ancestral Sequences. The aligned sequences were used to construct a plausible ancestral sequence based on parsimony considerations. The ancestral sequence obtained was such that, at each position, a minimum total base change was required for mutation to the residues at the corresponding positions, in all the sequences. This procedure was also used to deduce the ancestral sequences for monocot and dicot sequences separately. The number of base changes needed to convert any residue to a gap was assumed to be three. Plots of base change averaged over a window of specified length, placed at each residue position of the aligned set, were obtained from the ancestral sequence.

Hydrophobicity in the present case is best represented as the average over the aligned residues.

$\langle H_i \rangle = \sum_j H_{ij}/n$, where n is the number of sequences, j is the index over sequences, and i is the relative position of the residue in the sequence.

Disulphide Connectivity Diagram. The disulphide connectivity diagram for dicots was based on experimental results (Laskowski and Kato 1980). The diagram for monocots was prepared by assuming that the connectivity is similar to that of dicots.

Results

Primary Structure of Horsegram Inhibitor

The complete amino acid sequence of HGI-III (76 amino acid residues) is shown in Fig. 1. The amino terminal analysis of CM-HGI-III led to the unambiguous identification of residues 1–30. Digestion of PE-HGI-III with Lys C endopeptidase yielded four peptides. The amino-terminal sequence contained the peptides L-2, L-3, and a partial overlap of L-5, which completed the stretch of 1–62 (Asp^1–Met^{62}). The peptide CNBr-1 obtained by chemical cleavage and the Lys C peptides L-4, L-1, and T-3 obtained by limited TCPK-trypsin digestion of CM-HGI-III provided the carboxy-terminal stretch 62–76 and completed the sequence. The sequences of the peptides T-1, T-5, T-6, and CNBr-2, although repetitive, confirmed the sequences obtained earlier. Peptides CNBr-1, L-2, and T-3 confirmed the carboxy-terminal sequence which was determined by carboxypeptidase digestion to be . . . HDD. The molecular mass calculated on the basis of sequence analysis (MW 8,321 Da) compares well with that determined by electron spray mass spectrometry (MW 8,625 Da). This sequence was identical to that of *Macrotyloma axillare* DE-3 (DOLAX3), except at positions 4, 19, and 21 (Fig. 3).

Alignment Scores

Figure 2 (A–C) shows the alignment scores for all pairs of sequences. Among BBIs from dicots, the peanut inhibitors (ARAHY) show low alignment scores (<10.0) with others. Broad bean (VICFA) and common vetch (VICAN) show scores less than 14.0. The scores between other pairs are appreciably higher (Fig. 2A). A similar clustering pattern of dicot inhibitors was observed by Norioka and Ikenaka (1983). BBI sequences from different monocot species exhibit wider differences in comparison with those of dicots (Fig. 2B). The scores are mostly less than 14.0, except between the N-terminal half of barley (HORVUN) and wheat (WHEAT1), which has a score of 21.2. The C-terminal half of barley (HORVUC) is closer to the C-terminal half of rice inhibitor (ORYSAC). The alignment scores between the dicot and monocot BBI domains (Fig. 2C) were significantly lower than those presented in Fig. 2A and Fig. 2B.

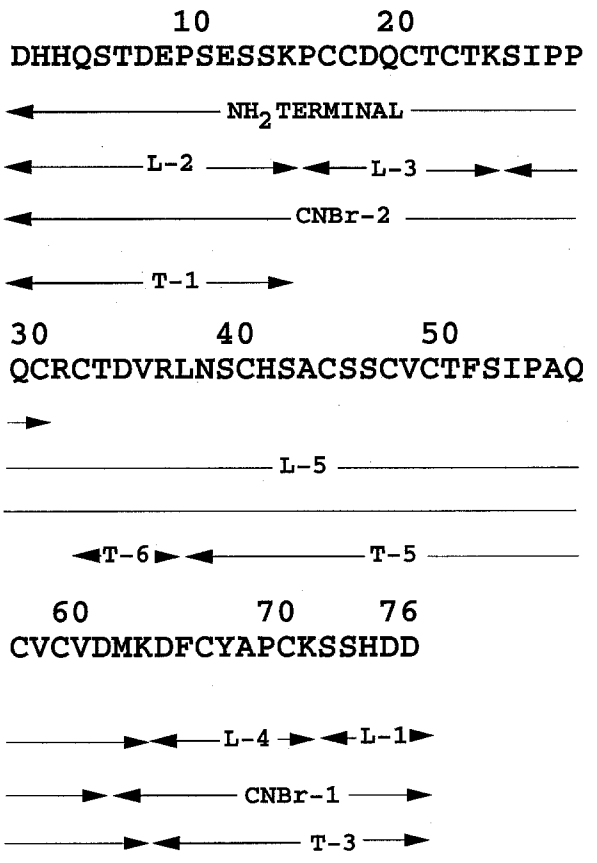

Fig. 1. Summary of primary structure determination of *Dolichos biflorus* (horsegram) 8K Bowman-Birk inhibitor. *T, L,* and *CNBr* denote trypsin, lysyl endopeptidase, and cyanogen bromide fragments, respectively.

Multiple Alignment: Variation in the Reactive Sites

The alignment of 27 domains of BBIs from both monocots and dicots is shown in Fig. 3. Dicot BBIs have 14 cysteins while the monocot BBIs of similar length have 10 cysteins (except ORYSAN, which has lost two more cysteins, Cys 4 and 5). The cysteins lost in monocots correspond to residues 3, 10, 11, and 13 of dicots (Fig. 4;

A	Arahy1	Arahy2	Medsa	Vican	Vicfa	Soybn	Phaau	Soybn3	Soybn2	Phaan2	Phavu	Hgi	Dolax3	Lonca4	Dolax4	Phalu	Phaan
Arahy1	25.4																
Arahy2	18.4	24.6															
Medsa	9.2	9.3	20.5														
Vican	7.7	7.2	13.2	23.9													
Vicfa	10.3	7.3	13.3	21.0	26.2												
Soybn	6.5	7.0	12.6	12.2	11.4	21.0											
Phaau	8.9	8.2	14.3	11.5	12.7	13.8	23.0										
Soybn3	8.6	7.0	11.7	9.1	12.0	12.0	16.6	28.2									
Soybn3	7.8	7.4	12.5	10.0	12.0	14.9	16.3	19.8	28.2								
Phaan2	8.9	7.7	10.7	11.5	10.8	14.3	17.5	13.9	15.5	27.1							
Phavu	9.0	7.2	13.6	10.8	10.9	12.4	17.5	16.6	17.5	16.8	25.7						
Hgi	7.4	7.5	13.3	11.3	11.8	15.2	16.0	13.6	16.4	20.4	16.3	21.8					
Dolax3	6.9	7.2	14.6	11.0	12.7	20.5	15.6	14.8	14.8	18.2	16.2	25.7	24.9				
Lonca4	5.9	4.7	11.3	11.5	10.6	14.1	15.3	13.0	15.1	17.0	16.0	16.3	16.7	23.1			
Dolax4	7.5	6.2	13.0	11.0	13.2	17.8	18.1	14.3	16.5	18.1	16.7	15.9	17.6	18.7	25.3		
Phalu	6.2	6.2	15.7	13.1	12.0	18.0	16.1	11.9	13.2	20.9	16.2	21.3	22.6	15.4	17.4	22.5	
Phaan	5.7	6.3	11.4	11.3	11.3	16.8	5.2	18.9	15.5	18.1	16.7	17.6	14.0	16.0	16.5	20.5	25.8

B	Wheat1	Horvun	Orysac	Setit2	Setit3	Coila1	Coila2	Horvuc	Orysan	Wheat2
Wheat1	23.1									
Horvun	21.2	25.4								
Orysac	11.9	12.2	22.5							
Setit2	10.8	13.0	13.2	25.6						
Setit3	10.1	13.0	14.9	23.9	24.7					
Coila1	11.4	13.8	12.2	11.0	15.3	25.4				
Coila2	11.4	13.8	12.2	11.0	15.3	25.4	25.4			
Horvuc	11.5	11.8	14.0	11.4	10.6	13.4	13.4	24.5		
Orysan	6.8	9.3	7.9	10.2	9.7	12.2	12.2	9.0	23.9	
Wheat2	15.3	15.4	10.7	12.5	13.3	11.9	11.9	11.8	9.0	20.9

C	Wheat1	Horvun	Orysac	Setit2	Setit3	Coila1	Coila2	Horvuc	Orysan	Wheat2
Arahy	5.5	6.2	8.3	6.3	5.5	5.3	5.3	6.3	2.3	4.5
Arahy2	3.6	4.5	6.0	5.5	5.2	4.0	4.0	6.9	2.5	3.8
Medsa	5.4	4.5	5.4	7.3	7.1	6.6	6.6	7.4	2.9	5.5
Vican	5.3	5.7	5.4	5.9	7.0	7.5	7.5	8.5	3.1	5.1
Vicfa	6.5	5.1	6.8	7.7	7.0	8.8	8.8	8.1	2.8	6.7
Soybn	5.8	4.8	5.6	6.7	5.9	6.1	6.1	6.6	4.6	5.5
Phaau	4.0	3.8	4.6	5.5	5.1	6.0	6.0	6.3	2.7	4.6
Soybn3	6.4	4.0	4.4	4.4	4.8	6.2	6.2	5.7	0.3	4.1
Soybn2	4.2	3.5	3.1	4.8	5.1	5.8	5.8	5.3	1.0	3.9
Phaan2	5.0	5.0	4.0	5.3	5.7	6.1	6.1	6.1	1.5	5.0
Phavu	5.2	5.4	3.8	5.0	4.8	6.6	6.6	7.0	1.6	4.6
Hgi	6.9	5.3	4.6	5.6	5.7	6.7	6.7	6.3	2.1	6.9
Dolax3	6.1	5.3	4.5	5.9	5.9	5.5	5.5	5.9	2.0	5.4
Lonca4	4.0	4.1	4.1	5.0	4.3	4.9	4.9	5.2	1.0	3.9
Dolax4	6.7	6.4	4.1	6.7	6.9	6.0	6.0	5.3	2.7	6.0
Phalu	6.3	6.0	4.4	7.4	6.5	6.4	6.4	7.9	2.8	6.0
Phaan	6.3	4.4	4.5	6.1	5.5	5.5	5.5	6.4	1.8	6.2

Fig. 2. Alignment scores for different pairs of Bowman-Birk inhibitor domains. Comparison between **(A)** dicot—dicot domains, **(B)** monocot—monocot domains, and **(C)** dicot—monocot domains.

cysteins are numbered 1–14 in the order of appearance along the dicot sequence).

Despite close similarity between their sequences, dicot BBIs show a broader specificity for the enzyme inhibited, when compared to monocot inhibitors. The first site of dicot inhibitors inhibits trypsin while the second site inhibits trypsin, chymotrypsin, or elastase. A closer look at the reactive-site segments of aligned sequences reveals the following features: The first site in most of the dicot inhibitors has residues Lys/Arg and Ser at the P1 and P1′ positions, respectively, and inhibits trypsin, except in soybean inhibitor C-II (SOYBN2) and kidney bean inhibitors (PHAVU) where the P1 position is occupied by Ala and hence the inhibition is specific to elastase.

In monocots, the only reactive site situated at the N-terminal region of the 8K protein aligns well with the first reactive site of the dicots. All monocots inhibit trypsin but seem to have lost their second reactive site. The residues corresponding to the P1 position of this site are

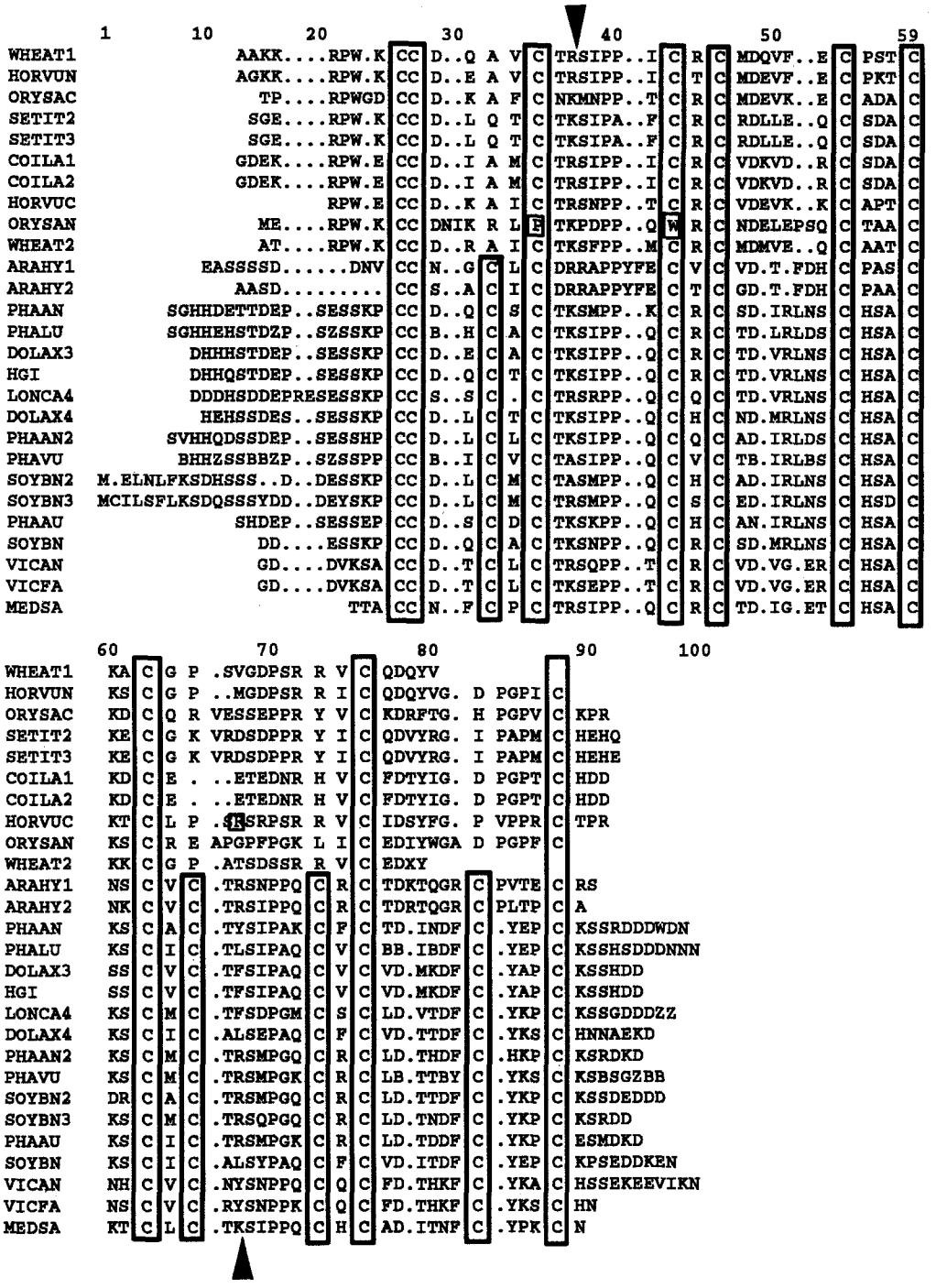

Fig. 3. Alignment of 26 Bowman-Birk inhibitor domains. Only cysteine residues have been *boxed* to highlight disulphide deletion in monocots. The *arrowheads* are between the P1 and P1′ residues of the reactive sites.

not the potential targets for trypsin, chymotrypsin, or elastase. More importantly, these inhibitors have lost Cys 10 and Cys 11, which form a disulphide linkage in dicots (Fig. 4). This disulphide holds the inhibitory loop and presents it to the enzyme in a favorable conformation (Chen et al. 1992; Tsunogae et al. 1986). The loss of this crucial disulphide and the lack of the appropriate P1 residue are probably the contributing factors for the loss of the second reactive site in monocots. The consensus sequence for the reactive sites of BBIs derived from Fig. 3 is shown in Fig. 5.

Hydrophobicity Patterns

The variation of hydrophobicity (Kyte and Doolittle 1982) along the polypeptide sequence is an important characteristic that often reflects the role of a given seg-

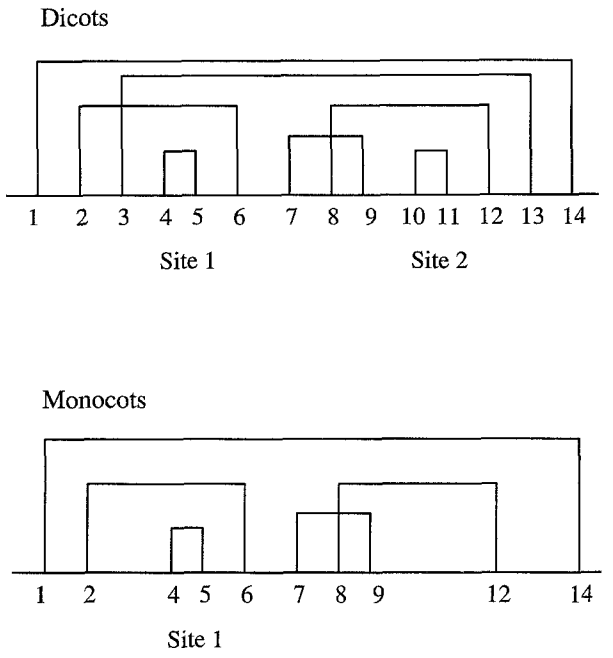

Fig. 4. Disulphide connectivity in dicot and monocot Bowman-Birk inhibitors.

Species	Reactive Site	Inhibition	Consensus Sequence								
			P3	P2	P1	P1'	P2'	P3'	P4'	P5'	P6'
Dicots	I	Total	C	T	*	S	X	P	P	Q	C
Monocots	I	Total	C	T	*	S	I	P	P	X	C
Dicots	I	Total	C	T	#	S	X	P	X	Q	C
Monocots	II	None	X	X	X	X	+	P	X	R	X
Orysan	I	Partial	P	T	K	P	D	P	P	Q	W
Horvuc	II	None	P	S	R	S	R	P	S	R	R
Dicots	I & II	] Total	C	T	#	S	X	P	P	Q	C
Monocots	I	]									

* -- K/R

-- K/R/F/Y/L

X -- Any residue.

+ -- Positvely charged residue.

Fig. 5. Consensus sequence of the inhibitory loops of Bowman-Birk inhibitors as obtained from the multiple alignment. *X* indicates any of the 20 residues; + indicates a positively charged residue.

ment in the native structure of the protein. Figure 6 shows hydrophobicity plots as a function of residue position, averaged over all the sequences, for dicots and monocots separately. The most hydrophilic regions of the polypeptide in dicot inhibitors occur toward the amino- and carboxy-terminal ends. These terminal residues do not have well-defined electron density in the X-ray structure of proteins from soybean (Chen et al. 1992) and peanut (Suzuki et al. 1993). The second site of monocots, although not active, is strongly hydrophilic. This increased hydrophilicity might have arisen due to the fact that the disulphide is lost in this segment, resulting in higher flexibility and water accessibility of this loop.

Phylogenetic Tree

Figure 7 shows the unrooted phylogenetic tree obtained from the PHYLIP suite of programs. The reliability of the branches was evaluated by a bootstrap procedure. The tree has two major branches. The circled region corresponds to the BBIs from monocots while the rest are BBIs from dicots. The tree gives a clear separation between BBIs from monocots and dicots.

Ancestral Sequences

Ancestral amino acid sequences were constructed for monocots and dicots separately (Fig. 8). The two ancestral sequences have 36% identity. The ancestral sequence corresponding to dicots has 45–82% identity with individual dicot sequences while it has 34–44% identity with monocot sequences. Similarly the monocot ancestral sequence has identities of 56–75% with monocot inhibitors and 30–45% with dicot inhibitors.

The evolutionary divergence of the sequences from the derived ancestral sequence was evaluated in terms of the minimum number of base changes required for mutating a residue in the ancestral sequence to each of the residues found in the sequences.

Figure 9 shows minimum base change as a function of residue position, for dicots and monocots, separately. In the plot, the reactive site regions are marked. The largest variation occurs at the amino- and carboxyl-terminal regions. The region corresponding to the reactive sites of dicots and the first reactive site of monocots show minimum change. In comparison, the residues of the second nonreactive site of monocots show a larger base change, in tune with the loss of function.

Discussion

BBI from horsegram (HGI-III) is a single polypeptide of 76 amino acid residues (MW ≈ 8,300 Da) (Fig. 1). Among BBIs, this sequence showed a maximum similarity to DE-3 from *Macrotyloma axillare* (DOLAX3). HGI has 76% and 41% identities with soybean and peanut inhibitors, respectively, for which the three-dimensional structures are known (Chen et al. 1992; Suzuki et al. 1993) (Fig. 3).

The sequence analysis presented in this paper suggests that the BBIs from monocot sources need to be grouped as a subclass of the BBI family of inhibitors. The phylogenetic tree constructed from all the available BBI sequences (Fig. 7) segregates monocot and dicot inhibitors. The amino-terminal BBI domain of RBTI (ORYSAN) shows anamously low alignment scores when compared to pairs of other inhibitor domains.

The three-dimensional structure of the dicot 8K inhibitors from soybean (Chen et al. 1992) and peanut (Suzuki et al. 1993) consists of two similar 4K domains

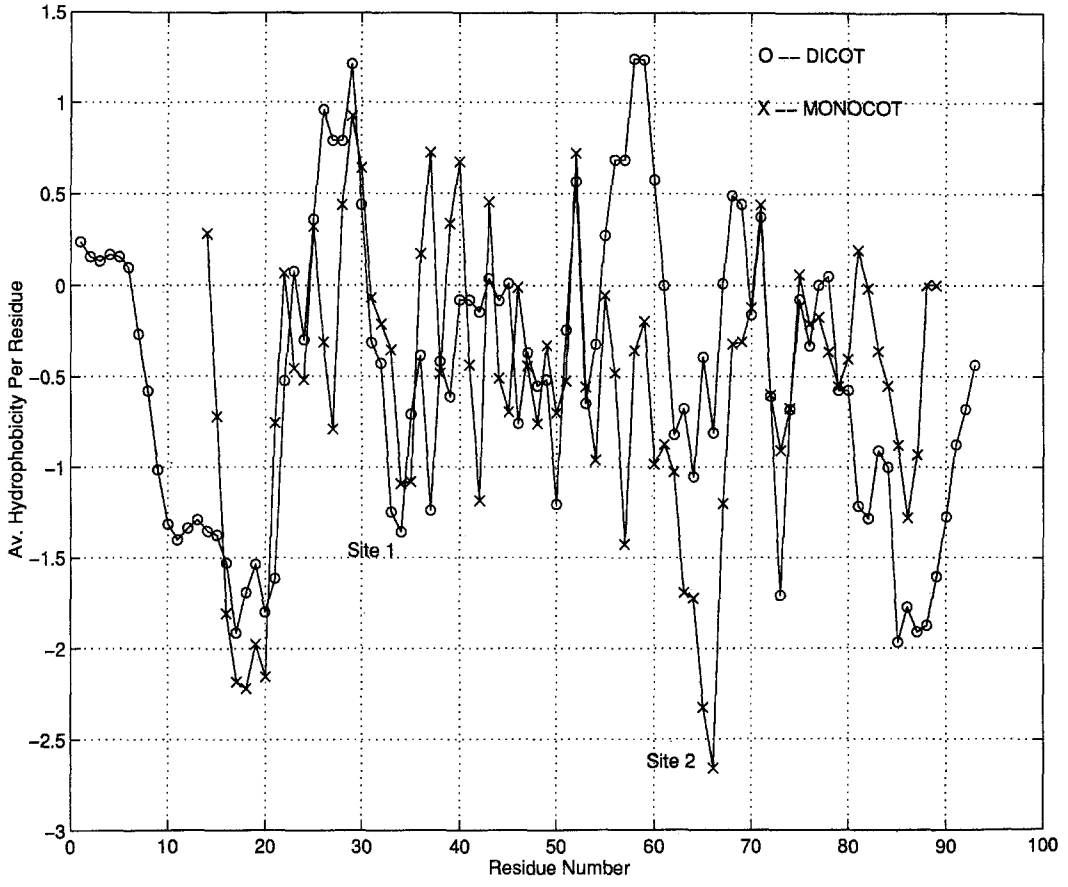

Fig. 6. Plot of hydrophobicity along the polypeptide averaged over the aligned inhibitor sequences. (○ for dicots; × for monocots). The two reactive sites, *site 1* and *site 2*, are marked. The plot corresponding to monocots has been translated suitably such that its active sites overlap with the active sites of dicots.

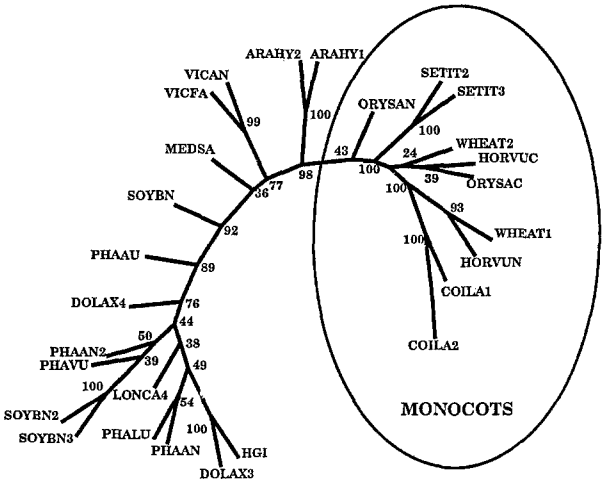

Fig. 7. Phylogeny of Bowman-Birk inhibitors. The *circled region* corresponds to inhibitors from monocotyledonous plants and the rest of the figure corresponds to those from dicotyledonous plants. The unrooted tree was produced as described in Methods. The *numbers* at the forks indicate the number of times the group consisting of the species which are to the right of that fork occurred among the trees out of 100 trees used to determine the consensus tree.

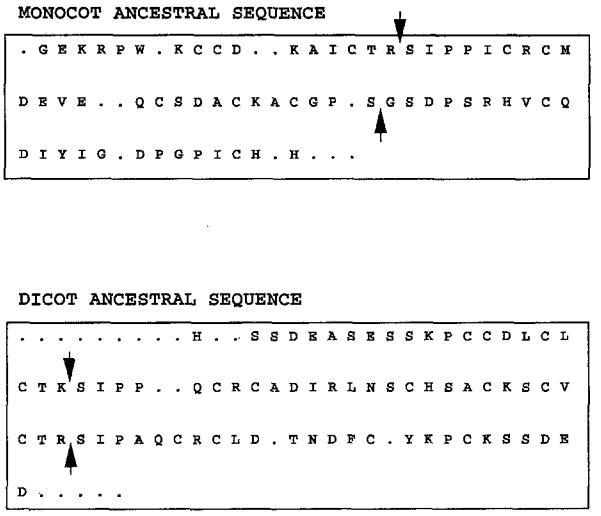

Fig. 8. The ancestral amino acid sequences of dicot and monocot inhibitors, obtained from the aligned sequences. The arrowheads are between P1 and P1′ residues of the reactive sites.

linked by the disulphide between Cys 3 and Cys 13. Accessible surface-area calculations for these two structures show that the solvent accessibility of most of the residues is greater than 50 Å^2 (data not shown). Hence, the BBI domains do not appear to have a strong hydrophobic core. The disulphide bonds seem to be the major contributors to structural stability of BBIs. Each domain of dicot BBIs has a loop responsible for inhibition. The

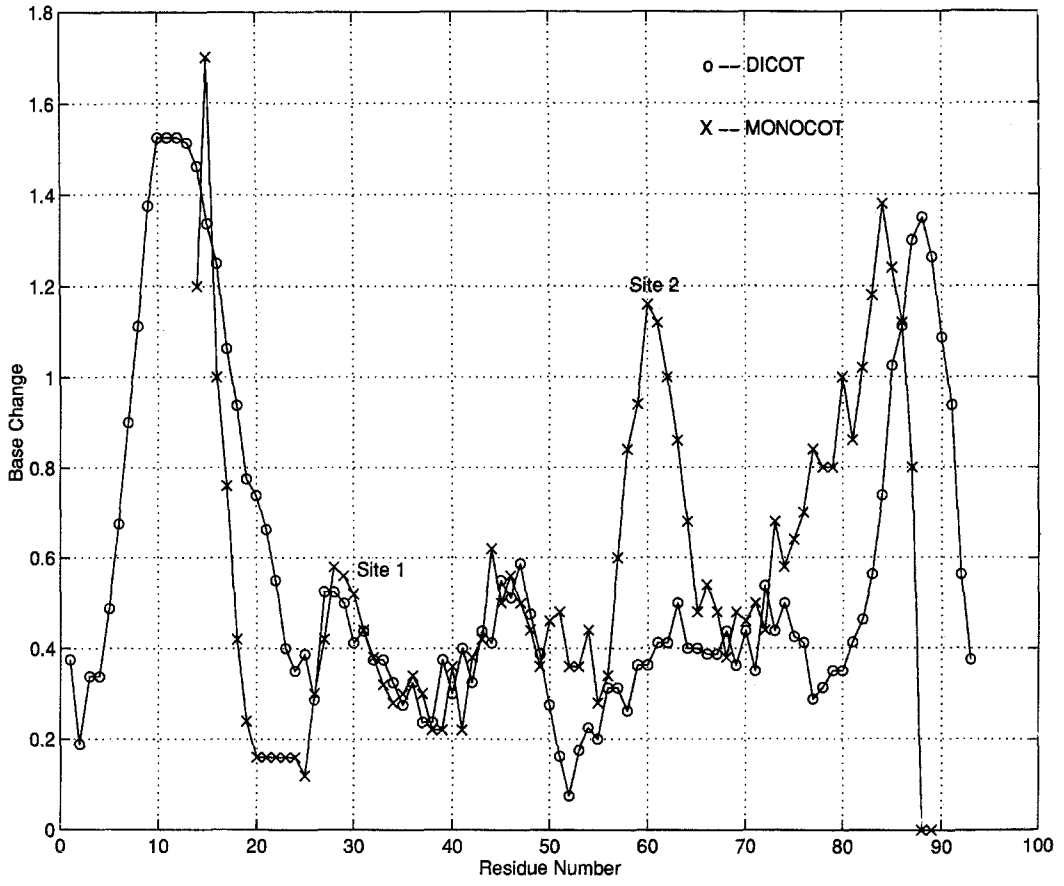

Fig. 9. Plot of minimum base changes required to convert five consecutive residues of the deduced ancestral sequence to all the residues found at the corresponding sequences (O for dicots; × for monocots). The two reactive sites, *site 1* and *site 2*, are marked. The plot corresponding to monocots has been translated suitably such that its active sites overlap with the active sites of dicots.

most accessible residues of the structure are the P1 residues of the two loops. In contrast, the monocot 8K inhibitors are single headed and inhibit only trypsin. They have lost two disulphides, viz. 10–11; 3–13 (Fig. 4). The functional implications of the absence of these disulphides were examined. One of these disulphides, 10–11, holds the second inhibitory loop in dicots. The absence of this disulphide increases the flexibility of the inhibitory loop. The average hydrophobicity plots (Fig. 6) show that this loop has acquired a larger hydrophilic character. The base change plots (Fig. 9) show that it also accumulates mutations at a higher rate. The loss of this disulphide seems to be correlated to the loss of second reactive site in monocots.

Rice BBI (RBTI) has lost two additional cysteins. These residues, Cys 4 and Cys 5, occur in the first site of the N-terminal 8K domain (ORYSAN) of this 16K protein (Fig. 5). In dicots, these cysteins form a disulphide bridge which holds the first inhibitory loop. In spite of the loss of this disulphide, RBTI retains activity and has a dissociation constant (K_D) which is two orders of magnitude greater than that of the other equivalent site in the C-terminal 8K fragment (Maki et al. 1980). Hence, the loss of the disulphide 4–5 does not lead to complete abolition of inhibitory activity. This suggests that the second site in monocots might also exhibit a low level of inhibition. It is observed that the P1 residue is an Arg (Highlighted in Fig. 3) in the second site at the carboxy-terminal half of the barley inhibitor (HORVUC). However, this site is not known to have any activity. Therefore a potential P1 residue is perhaps not sufficient for a site to show inhibition. The consensus sequence for the various reactive sites of dicots and monocots, shown in Fig. 5, is CTXSXPPQC. The conserved residues appear to interact with the residues lining the active site pocket of trypsin (Lin et al. 1993). In rice, six out of nine loop residues (P2, P1, P1′, P3′, P4′, P5′) are conserved, whereas in barley only three (P1, P1′, P3′) are conserved. However, only the P3′ residue is conserved at the second site of monocots. Hence, the conservation of six residues of the N-terminal site of rice without the disulphide bridge seems to be sufficient for its partial activity. The somewhat lower alignment scores of ORYSAN (Fig. 2C) when compared to other pairs of inhibitors is probably due to the loss of this disulphide. Determination of the three-dimensional structure of the monocot inhibitors will provide further insight into the structural consequences of these mutations.

The present analysis of BBI sequences from monocots and dicots shows that monocot 8K inhibitors have lost their C-terminal reactive site during the process of evolution (Tashiro et al. 1990). It is likely that gene duplication in monocots leading to the evolution of 16K double-headed BBI occurred after the loss of the second site in the monocot 8K proteins. Probably due to the loss of second site, monocot BBIs show greater variability in sequence when compared to the dicots. These observations also suggest that gene duplication occurred only in monocots and that dicots are unlikely to have a 16K inhibitor. However, we had earlier reported preliminary X-ray diffraction studies on a 16K BBI from a dicot, horsegram (*Dolichos biflorus* or *Macrotyloma uniflorum*) (Prakash et al. 1994). This estimate of molecular weight was based on gel filtration and SDS-PAGE in the presence of reducing agents (Rama Sarma and Rajagopal Rao 1991). The molecular weight of horsegram inhibitor has been redetermined by mass spectrometric analysis (Gowda, Sreerama, and Rajagopal Rao, private communications) and sequencing. In conformity with the present analysis, these studies show that the BBI from horsegram is an 8K protein and not a 16K protein as reported earlier.

Acknowledgments. S.S. thanks the Department of Biotechnology, Government of India, and the authorities of Bharathidasan University, Tiruchirapalli, for financial support. M.R.N.M. thanks the Department of Science and Technology for financial support. Y.N.S. thanks CSIR, New Delhi, for financial assistance. Help from the Bioinformatics Center, Indian Institute of Science, Bangalore; the National Sequencing Facility at the Department of Biochemistry, Indian Institute of Science, Bangalore; and from Dr. H.S. Savithri, Ms. M. Bhuvaneshwari, and Mr. Sameer Velankar is gratefully acknowledged.

References

Ambler RP (1972) Enzymic Hydrolysis with carboxypeptidases. Methods Enzymol 25:143–154

Birk Y (1985) The Bowman-Birk inhibitor. Int J Pept Protein Res 25:113–131

Bidlingmeyer BA, Cohen SA, Tarvin TL (1984) Rapid analysis of amino acids using pre-column derivatization. J Chromatogr 336: 93–104

Chang JY (1979) The destruction of serine and threonine thiohydantoins during the sequence determination of peptides by 4-N,N-dimethylaminoazobenzene 4′-isothiocyanate. Biochem Biophys Acta 578:175–187

Chen P, Rose J, Love R, Wei CH, Wang BC (1992) Reactive sites of an anticarcenogenic Bowman-Birk Protease inhibitor are similar to other trypsin inhibitors. J Biol Chem 267:1990–1994

Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic acids Res 16:10881–10890

Crestfield AM, Moore S, Stein WH (1963) The preparation and enzymatic hydrolysis of reduced and S-carboxymethylated proteins. J Biol Chem 238:622–627

Desai N, Bourne P (1986) Protein and nucleic acid sequence information and analysis, PRONUC ver 4.4. Columbia University, New York

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet 22:521–565

Felenstein J (1989) PHYLIP—Phylogeny inference package (version 3.2). Cladistics 5:164–166

Hermodson MA, Erickson LH, Neurath H, Walsh KA (1973) Determination of the amino acid sequence of porcine trypsin in sequenator analysis. Biochemistry 12:3146–3153

Joubert FJ, Kruger H, Townshend GS, Botes DP (1979) Purification, some properties and the complete primary structures of two protease inhibitors (DE-3 and DE-4) from *Macrotyloma axillare* seed. Eur J Biochem 97:85–91

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 151:105–132

Laskowski M Jr, Kato I (1980) Protein inhibitors of proteinases. Annu Rev Biochem 49:593–626

Lin G, Bode W, Huber R, Chi C, Engh RA (1993) The 0.25 nm X-ray structure of the Bowman-Birk inhibitor from mung bean in ternary complex with porcine trypsin. Eur J Biochem 212:549–555

Lipman DJ, Pearson WR (1985) Rapid and sensitive protein similarity searches. Science 227:1435–1441

Mahoney WC, Hermodson MA (1980) Separation of large denatured peptides by reverse phase high performance liquid chromatography. J Biol Chem 255:11199–11203

Maki Z, Tashiro M, Sugihara N, Kanamori M (1980) Double headed nature of a trypsin inhibitor from rice bran. Agric Biol Chem 44: 953–955

Nagasue A, Fukamachi H, Ikenaga H, Funatsu G (1988) The amino acid sequence of barley rootlet trypsin inhibitor. Agric Biol Chem 52:1505–1514

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Norioka S, Ikenaka T (1983) Amino acid sequences of trypsin chymotrypsin inhibitors (A-I,A- II,B-I and B-II) from peanut (*Arachis hypogaea*): a discussion on molecular evolution of legume Bowman-Birk type inhibitors. J Biochem 94:589–599

Odani S, Koide T, Ono T (1986) Wheat germ trypsin inhibitors. Isolation and structural characterization of single-headed and double-headed inhibitors of the Bowman-Birk type. J Biochem 100:975–983

Prakash B, Murthy MRN, Sreerama YN, Rama Sarma PR, Rajagopal Rao D (1994) Crystallization and preliminary X-ray diffraction studies on a trypsin/chymotrypsin double headed inhibitor from horse gram. J Mol Biol 235:364–366

Rama Sarma PR, Rajagopal Rao D (1991) Nature of tryptic /chymotryptic inhibitor from horse gram (*Dolichos biflorus*) Indian. J Biochem Biophys 28:418–424

Ryan CA (1981) Proteinase inhibitors, a comprehensive treatise. In: Stumpf PK, Conn EE (eds) The Biochemistry of Plants, vol 6. Academic Press, New York, p 351–371

Schwartz RM, Dayhoff MO (1978) Matrices for detecting distance relationships. In: Dayhoff MO (ed) Atlas of protein sequence and structure. Washington DC. National Biomedical Research Foundation, 5:353–358

Suzuki A, Yamane T, Ashida T, Norioka S, Hara S, Ikenaka T (1993) Crystallographic refinement of Bowman-Birk type protease inhibi-

tor A-II from peanut (*Arachis hypogaea*) at 2.3 Å resolution. J Mol Biol 234:722–734

Tashiro M, Hashino K, Shiozaki M, Ibuki F, Maki Z (1987) The complete amino acid sequence of rice bran trypsin inhibitor. J Biochem 102:297–306

Tashiro M, Asao T, Hirata C, Takahashi K, Kanamori M (1990) The complete amino acid sequence of a major trypsin inhibitor from seeds of foxtail millet (*Setaria italica*). J Biochem 108:669–672

Tsunogae Y, Tanaka I, Yamane T, Kikkawa J, Ashida T, Ishikawa C, Watanabe K, Nakamura S, Takahashi K (1986) Structure of trypsin binding domain of Bowman-Birk type protease inhibitor and its interaction with trypsin. J Biochem 100:1637–1646

Waxadal MJ, Konigsberg WH, Henley WL, Edelman GM (1968) The covalent structure of a human γg-immunoglobulin—II. Isolation and characterization of the cyanogen bromide fragments. Biochemistry 7:1959–1966

Yang CV, Paulay E, Krazer HS, Hilschmann N (1981) Chromatography and rechromatography in high performance liquid chromatography of peptide mixtures: the complete primary structure of an immunoglobulin L-chain of k-type, subgroup I (Bence-Jones protein Den). Hoppe Seylers Z Physiol Chem 362:1131–1146