Isolation and functional diversity of Bowman-Birk type serine proteinase inhibitors from 

*Hyacinthus orientalis*

Narumi Aoki-Shioi†,*,1, Shigeyuki Terada†,1, Roland Hellinger2, Yoshitaka Furuta1, Christian W. Gruber2

1Department of Chemistry, Faculty of Science, Fukuoka University, 19-1, 8-chome Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan

2Center for Pharmacology and Physiology, Medical University of Vienna, Schwarzspanierstr. 17, 1090 Vienna, Austria

†These authors contributed equally to this work.

*To whom correspondence should be addressed: Narumi Shioi: Department of Chemistry, Faculty of Science, Fukuoka University, 19-1, 8-chome Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan; anarumi@fukuoka-u.ac.jp; Tel. +81-(0)92-870-6631 ext. 6215; Fax.+81-(0)92-865-6030;

Narumi Shioi: anarumi@fukuoka-u.ac.jp

Shigeyuki Terada: sterada@fukuoka-u.ac.jp

Roland Hellinger: roland.hellinger@meduniwien.ac.at

Yoshitaka Furuta: ange-boucle.@docomo.ne.jp

Christian W. Gruber: christian.w.gruber@meduniwien.ac.at

Running title: The diversity of monocot serine proteinase inhibitors

Keywords: plant peptide, cysteine-rich, serine proteinase inhibitor, lily, precursor gene
Abstract

Bowman–Birk inhibitors (BBI) are plant-derived serine proteinase inhibitors. Endogenously, they function as defense molecules against pathogens and insects, but they also have been explored for applications in cancer treatment and inflammatory disorders. Here, we isolated 15 novel BBIs from the bulb of *Hyacinthus orientalis* (termed HOSPIs). These isoinhibitors consisted of two or three chains, respectively, that are linked by disulfides bonds based on proposed cleavage sites in the canonical BBI reactive site loop. They strongly inhibited trypsin ($K_i = 0.22 - 167 \text{ nM}$) and $\alpha$-chymotrypsin ($K_i = 19 - 1200 \text{ nM}$). Notably, HOSPI-B4 contains a six-residue reactive loop, which appears to be the smallest such motif discovered in BBIs to date. HOSPI-A6 and -A7 contain an unusual reactive site, i.e. Leu-Met at the P1-P1’ position and have strong inhibitory activity against trypsin, $\alpha$-chymotrypsin and elastase. Analysis of the cDNA encoding *HOSPIs* revealed that the precursors have HOSPI-like domains repeated at least twice with a defined linker sequence connecting individual domains. Lastly, mutational analysis of HOSPIs suggested that the linker sequence does not affect the inhibitory activity, and a Thr residue at the P2 site and a Pro at the P3’ site are crucial for elastase inhibition. Using mammalian proteases as representative model system, we gain novel insight into the sequence diversity and proteolytic activity of plant BBI. These results may aid the rational design of BBI peptides with potent and distinct inhibitory activity against human, pathogen, or insect serine proteinases.

Introduction

Proteinase inhibitors (PI) are widely distributed in all domains of life and they are involved in the regulation of several biochemical processes. For instance, in animals PIs are important for blood coagulation regulation, immune system activation, and to tune the activity of proteinase processing of peptide hormones [1]. In plants, PIs are involved in maintaining physiological homeostasis and serve the plants’ innate defense machinery against herbivore attacks or microbial infections [2]. Research on PIs in plants started with the discovery of a soybean trypsin inhibitor by Kunitz in 1947 [3]. Since
then, many plant-derived small molecule and proteinaceous PIs have been discovered. Currently there are more than 6700 individual plant PIs known that are classified into at least 12 distinct types, such as serpins, phytocystatins, Kunitz inhibitors, Bowman-Birk-, α-amylase-trypsin-, mustard-type-, potato type-I-, potato type-II-, potato metallocarboxypeptidase-, squash- and cyclotide inhibitors [2, 4]. Therefore these peptides PIs provide promising molecules for discovery and development of drug candidates or agrochemicals.

Bowman-Birk inhibitors (BBI) from soybean have been studied in clinical trials as anticancer agents for melanoma and oral cancers [5, 6]. BBIs are cysteine-rich proteins typically consisting of 60-90 amino acids including 10-12 cysteines, which form a cysteine framework. BBI stability is a result of the disulfide bonds supporting a compact protein structure and a rigid conformation, which result into a considerable resistance against heat, acid and basic conditions (Figure S1A-C) [7-9]. They were originally isolated from species of the legume family of dicotyledonous plant [10]. The inhibitory domain (BBI domain) consists of an anti-parallel β-sheet motif with two inhibiting loops on opposite sides of the molecule [2, 9], which is also referred to as double-headed inhibitor (Figure S1B) [11]. BBIs have also been isolated from monocotyledonous plants of the Poaceae family. Intriguingly these BBIs contain only one functional reactive site located on inhibition loop in one fully functional motif, whereas the lack of key disulfide bonds in the opposite face result in a loss of proteinase inhibitory capacity in the second motif (Figure S1A) [12, 13]. The canonical reactive sites of BBIs are present in the stabilized inhibition loops by highly conserved internal disulfide bond, which contain 9 to 11 residues. The enzyme specificity of BBIs is linked to the sequence of the reactive site; in particular the P1 site appears to be important to guide substrate preference of the target proteinase [14, 15].

We previously reported the amino acid sequences, gene structures and biochemical properties of BBIs from onion bulb (Allium cepa) as the first BBI found in Liliaceae spp. [16], which exhibit high sequence homology to leguminous plant PIs. In the onion-derived inhibitors, a particular cysteine residue near the second reactive site was absent leading to functional impairment of the second inhibitory module. However, we demonstrated that the onion BBI had not only in trypsin inhibition, but also in the inhibition of chymotrypsin [16]. Thus, the mechanism and selectivity of some BBIs
could not be rationally explained by consideration of the canonical reactive site (P1 and P1’ positions) only.

In this study, 15 novel proteinase inhibitors from hyacinth bulb (*Hyacinthus orientalis* L., formerly known to belong to the Liliaceae plant family) were isolated and their primary structures were determined. Based on homology analysis they belong to the family of BBI and display a characteristic highly conserved cysteine residues. Since there is limited information of monocot plant PIs, we were interested to investigate the diversity of BBI in hyacinth. Interestingly, these novel inhibitors had unique sequence motifs in their inhibition loop. Therefore, we determined their inhibitory activity against representative mammalian serine proteinases. Accordingly, they were termed *Hyacinthus orientalis* serine proteinase inhibitors, HOSPIs. To gain valuable insight into the sequence-activity relationship of these interesting peptide PIs, a cDNA transcript analysis was performed. Cloning of six *HOSPI* genes from the bulb revealed two distinct types of genes. Interestingly, the HOSPI-like domains were repeated at least twice in each gene with a signature linker sequence connecting the domains. To investigate the role of this linker motif, their sequence variations and the function of the different domains of the precursor, we established the expression system in *E. coli* of a fusion peptide with the linker and the mature HOSPI sequence derived from a representative HOSPI gene. These results suggested that the linker sequence does not affect the inhibitory activity. Further, a Thr residue at the P2 site and a Pro at the P3’ site are crucial for elastase inhibition. Thus, the surrounding residues of the reactive site might contribute to the recognition and/or interaction with the enzymes. In this study we are able to provide a system to study the functional variety of plant BBIs for the inhibition mechanism of mammalian serine proteinases.

**Materials and methods**

**Materials and reagents**

*Hyacinthus orientalis* L. bulbs were purchased at Home Center & Home Fashion Store, Nafco (Fukuoka, Japan). Proteinases were from Sigma Chem. Co. (St. Louis) and proteinase substrates (as outlined below) were purchased from the PEPTIDE INSTITUTE (Osaka, Japan).
Preparation of hyacinth extract and purification of inhibitors via chromatography

H. orientalis bulbs (626.2 g) were homogenized using a common household blender. The processed mixture was then centrifuged at 9000 rpm (KUBOTA 6500, KUBOTA Co., Tokyo, Japan) at 4°C for 40 min to obtain the aqueous supernatant containing the inhibitors of interest. The extract solution was thermally inactivated at 60°C for 30 min to remove precipitate proteins and as inactivation of enzymatic activity. The supernatant was mixed with saturated ammonium sulfate until made up final concentration 80%. The obtained precipitate was dissolved in 120 ml water, dialyzed and lyophilized to yield the crude inhibitor cocktail. Afterwards 2.0 g of crude inhibitor powder were dissolved in 50 mM Tris-HCl (pH 8.0) and applied to DEAE-Toyopearl 650 M ion exchange column (5.0 × 27 cm, TOSOH Co., Tokyo, Japan). The proteins were eluted with a gradient from 0 to 0.35 M NaCl in 50 mM Tris-HCl (pH 8.0). Finally, the proteinase inhibitors of each fraction were purified by reversed-phase high performance liquid chromatography (HPLC) using a YMC-Pack C8 column (1.0 x 15 cm, YMC CO. LTD., Kyoto, Japan) with a linear gradient of solvents 0.1% TFA (eluent A) and acetonitrile in 0.1% TFA (eluent B) at the flow rate of 3.0 mL/min. Protein elution was monitored by absorbance at 230 nm.

Mass spectrometry

Molecular mass of each purified inhibitor was analyzed by a Voyager DE-STR matrix-assisted laser desorption ionization time-of-flight (MALDI–TOF) mass spectrometer (Applied Biosystems Japan, Tokyo, Japan). The analysis was performed by following the manufacturer’s instructions. Briefly, approximately 3 µg of each sample was dissolved in 0.1% TFA-50% acetonitrile containing saturated aqueous solution of sinapinic acid (Sigma Chem. Co., St. Louis) and spotted for analysis onto the target plate.

Quantification of proteins
The concentration of each purified inhibitor was determined using the molar extinction coefficients \((\varepsilon \text{ M}^{-1}\text{cm}^{-1})\) at 280 nm calculated by ProtParam tool on ExPASy server [17]. The \(\varepsilon\) value of each inhibitor was listed in Supplementary Table S1.

**Proteinase inhibitory activity**

**Trypsin inhibitory activity:** 50 µL of trypsin dissolved in 50 mM Tris-HCl (pH 8.0) (final concentration 0.08 µM) and 100 µL of HOSPI dissolved with water (final concentrations 0.017, 0.033, 0.05, 0.067, 0.1 and 0.2 µM) were added to 1300 µL of 20 mM CaCl\(_2\) in 50 mM Tris-HCl (pH 8.0) and allowed to incubate at room temperature for 3 min. Then, 7.5 mM benzoyl-L-arginine \(p\)-nitroanilide (L-BAPA) dissolved in 50 µL of methanol (final concentration 0.25 mM) was added to the pre-incubated solution containing trypsin and HOSPI.

**Chymotrypsin inhibitory activity:** 50 µL of \(\alpha\)-chymotrypsin (final concentration 0.45 µM) and 100 µL of HOSPIs (final concentrations 0.001, 0.003, 0.005, 0.006, 0.01 and 0.02 µM) were added to 1300 µL of 20 mM CaCl\(_2\) in 50 mM Tris-HCl (pH 8.0) and allowed to incubate at room temperature for 3 min. Then, 50 µL of 4.5 mM benzoyl-L-tyrosine \(p\)-nitroanilide (Bz-Tyr-pNA), dissolved in dimethyl sulfoxide, DMSO) was added (final concentration 0.15 mM).

**Elastase inhibitory activity:** 50 µL of elastase (final concentration 0.05 µM) and 100 µL of HOSPI (final concentrations 0.005, 0.0015, 0.025, 0.038, 0.05 and 0.1 µM) were mixed with 0.1 M KCl in 0.01 M Tris-HCl (pH 8.0) and allowed to incubate at room temperature for 3 min. Then, succinyl-L-alanyl-L-alanyl-L-alanine \(p\)-nitroanilide (STANA) dissolved in DMSO (final concentration 0.25 mM) was added to the pre-incubated solution as a substrate.

**Subtilisin inhibitory activity:** 50 µL of subtilisin (final concentration 0.25 µM) and 100 µL of HOSPI (final concentration 0.01, 0.10, and 0.20 µM) were added to in 50 mM CaCl\(_2\) in 0.05 M Tris-HCl, which was incubated at room temperature for 3 min. Then, 50 µL of 1.5 mM benzoyloxycarbonyl-L-alanyl-L-alanyl-L-leucine \(p\)-nitroanilide was added (final concentration 0.05 mM).

For each proteinase activity assay, the increase in absorbance at a wavelength of 410 nm was monitored to calculate the activity value. The apparent inhibition constant \(K_{\text{app}}\) was determined using...
a computational tool for curve fitting based on Morrison equation for tight binding of competitive inhibitors as shown below [18, 19].

\[
\frac{v_i}{v_0} = \frac{([E]+[I]+K_i^{app}) - \sqrt{([E]+[I]+K_i^{app})^2 - 4[E][I]}}{2[E]}
\]

The inhibition constant \( K_i \) was calculated from the following equation [20].

\[
K_i^{app} = K_i \left( 1 + [S]/K_m \right)
\]

\( K_m \) values for trypsin, chymotrypsin, and elastase were determined experimentally using the Lineweaver-Birk plot [21]. Final concentrations of each enzyme and the substrate were 0.08 \( \mu \)M bovine trypsin and between 0.2 - 0.5 mM L-BAPA, 0.45 \( \mu \)M bovine \( \alpha \)-chymotrypsin and 0.15 - 0.60 \( \mu \)M Bz-Tyr-pNA, 0.005 \( \mu \)M bovine elastase and 0.075 - 0.01 \( \mu \)M STANA for calculating \( -1/K_m \) of the plot. The \( K_m \) values for the trypsin, \( \alpha \)-chymotrypsin and elastase were 0.51, 2.04, and 0.423 mM, respectively.

**SDS-polyacrylamide gel electrophoresis (SDS-PAGE)**

SDS–PAGE was carried out on 12.5% gels by Laemmli’s method [22]. Prestained XL-ladder marker kit of Apro Science (Tokyo, Japan) was used for the estimation of molecular weights. The sample after separation on SDS-PAGE gel was fixed with trichloroacetic acid/methanol/water (1:4:5, w/v/v). Protein staining was performed using 0.25% Coomassie brilliant blue R-250 in methanol/acetic acid/water (4:1:5, v/v/v), and destaining was performed with methanol/acetic acid/water (0.1:1:8.9, v/v/v).

**Determination of amino acid sequence**

Each purified LOSPI (0.5-1.0 mg) was dissolved in 6 M guanidine hydrochloride, 10 mM EDTA, 0.5 M Tris-HCl (pH 8.6), and disulfide bonds were reduced with dithiothreitol (DTT) of 10 molar
equivalents against the number of cysteine residues. Then, reduced-HOSPI was allowed to react with 4-vinylpyridine of 3 molar equivalent for DTT at room temperature for 90 min to yield S-pyridylethylated proteins according to the procedure indicated in the reference [23]. By-product was eliminated by re-chromatography, and the modified proteins were purified. In this procedure, two or three peptide chains were detected in a separate by RP-HPLC after reductive reaction. The pyridylethylated protein was digested with various proteinases (Arg-C, V8 proteinase, Asp-N), and fragmented peptides were separated by reverse phase HPLC using a YMC-Pack ODS C18 column (4.6 x 250 mm) column. The amino acid sequence of each peptide was determined by N-terminal Edman degradation using a protein automated sequencer PPSQ series (Shimadzu, Kyoto, Japan).

cDNA cloning and expression of HOSPI by Escherichia coli

The total RNA was extracted from the bulbs of H. orientalis according to a standard protocol [24]. The cDNA encoding HOSPI was cloned using a method of Deshimaru et al. [16]. Briefly, cDNA was synthesized by reverse transcription from RNA using an adaptor-linked oligo (dT) primer (5'-GGCCACGCGTCGACTAGTAC(T)17'-3'). They were amplified by PCR using primers designed based on amino acid sequence of the purified natural inhibitor (Table S2). The primers LOSPI5-N (5'-TGYTGYAAYGARTGYGGNATHTGYGAY-3') and LOSPI6-N (5'-NTGYTGYAAYGARTGYGGNATHTGYGAY-3') were designed to the N-terminal region of LOSPI-B5 and -B6 (the amino acid sequence of Cys5 - Asp13), respectively, which was used in combination with the 3'-Adp primer (5'-GGCCACGCGTCGACTAGTAC-3'). Based on the resulting partial nucleotide sequences, a second primer (5'-TTGTGCCAATGCGCCACGAAGGCTGCC-3') corresponding to the amino acid sequence, Val-Pro-Gln-His-Glu-Gly-Cys, was designed based on the first PCR products. This sequence was observed in native inhibitors, LOSPI-A4, -B1, -B4, -B5, -B6 and -C1. Consequently, the partial sequences including the 3'UTR with poly A were determined for six clones by PCR using a combination of the second primer and the 3'-Adp primer.

After cDNA cloning of the partial LOSPI genes, we designed the following primers for cloning of the expression system: forward primers Ex-IkN (5'-
TATAACTGCGCCATGGACGGACATCGT-3’ and Ex-matN (5’-TTCAGAAGTGGCAGTTACCATGGCGTGG-3’), as well as the reverse primer Ex-matC (5’-TGCAAGCTTTCCCCAAGCGGCAGTTATA-3’). Then, the gene incorporated into expression vector pET32a through according procedure of pET expression system (Novagen, Darmstadt, HE). Nucleotide sequences of the expression vector were analyzed with ABI PRISM™ 377 DNA Sequencing System (Applied Biosystems, Forster City, CA), cDNAs having linker sequences (LOSPI-lk) and cDNAs without linker sequences (LOSPI-mat) were amplified by PCR and integrated into the expression vector pET32a. This was transformed into E. coli strain Rosetta-gamiB (DE3)pLysS (Novagen, Darmstadt, HE), and cultured until a density of OD600 of 0.6-0.7, and 1 mM isopropyl β-D-thiogalactoside was added to induce expression. The bacteria were collected and lysed by the BugBuster Protein Extract Reagent (Merck, Darmstadt, HE). The histidine-tagged recombinant proteins were purified on 1 mL Ni²⁺ Sepharose 6 Fast Flow resin in a micro-column (GE Healthcare, Chicago, IL) using by the binding buffer (50 mM Tris-HCl, 500 mM NaCl, 30 mM imidazole, pH 7.5) and elution buffer (50 mM Tris-HCl, 500 mM NaCl, 200 mM imidazole). The binding protein on the affinity chromatography was performed with reversed-phase HPLC for final purification on a same condition with native inhibitor described above.

Results

Isolation, purification and amino acid sequencing of LOSPIs

The aqueous extract from hyacinth bulb was applied to anion exchange chromatography (DEAE-Toyopearl), and each fraction was tested for inhibitory activity against trypsin (Figure 1A). The active fractions were pooled in three fractions (named A, B and C), which were carried forward to further isolation and characterization. The isolated proteins responsible for the trypsin inhibition were named H. orientalis serine proteinase inhibitors, HOSPIs. The seven inhibitors in fraction A (HOSPI-A1 to -A7, Figure 1B), seven inhibitors in fraction B (HOSPI-B1 to -B7, Figure 1C), and one inhibitor in fraction C (HOSPI-C1, Figure 1D) were isolated using by reversed-phase HPLC.
Amino acid sequences of proteinase inhibitors were determined by Edman degradation. As shown in Figure S2, full length sequences were obtained for HOSPI-A4, -A6, -A7, -B1, -B4 to -B7, and -C1, and partial sequences for HOSPI-A1 to -A3, -A5 and -B3 (Figure S2). Most but not all of these novel inhibitors were homologous to soybean BBI, which is a prototypical dicotyledonous plant BBI [25]. Sequence comparisons exhibited that HOSPIs only share low sequence homology with BBIs (18-29% identity), however, the cysteine residues in these inhibitors are well conserved as well as the typical BBI domain (Figure 2, Figure S1C). One of the disulfide bonds, usually embracing the second reactive site of soybean BBI, disappeared in HOSPIs. In a comparative analysis of amino acid sequences of HOSPIs, they were classified into two types: the 10 cysteine residues type and the 12 cysteine residues type. In addition, there are those of the N-terminus Gly (or Glu), or Trp-Pro-Pro-Val-Glu in which additional N-terminal extension are observed such as in HOSPI-A3 to -A7 and -B3 (Figure 2).

Seven HOSPIs (HOSPI-A2, -A3, -A5, -B4, -B5, and -B6) consist of two chains, and HOSPI-B1 has three chains (Table S1). The chains were identified by peptide sequencing of HPLC-derived S-pyridylethylated proteins after reduction of sulfhydryl group (refer to “P; the intact pyridylethylated protein before digestion by enzymes in Figure S2). These results suggested that the two or three chains were cross-linked by disulfide bonds. The amino acid sequences in the C-terminal region of these seven HOSPIs showed, Gly-Lys (Met or Thr)-Pro (His)-Leu (Met)-Tyr-Gln, corresponding to the second reactive loop in soybean BBI (Figure 2). Another cleavage site observed in HOSPI-B1 and -B4 was detected between Arg and Ser on a first reactive loop. The amino acid sequences between HOSPI-B4 and HOSPI-B1 shows high homology, which can be distinguished by a loss/insert of only two residues (Leu and Ser) in cleavage site of the C-terminal region described above. However, HOSPI-B4 has two peptide chains without the cleavage site in the C-terminal region. The first reactive site loop of HOSPI-B1 and -B4 are made up of only 6 residues (Cys-Ala-Arg-Ser-Leu-Cys), which is the smallest reactive loop in BBI reported so far. In summary, we isolated 15 novel proteinase inhibitors that belong to the BBI family.

Inhibitory activity of HOSPIs towards representative mammalian serine proteases
We next focused on investigating the nine HOSPIs with a determined full length amino acid sequence. To investigate the inhibitory activity ($K_i$) of the nine HOSPIs against serine proteinase, we used mammalian trypsin, chymotrypsin, and elastase and bacteria subtilisin as representative proteinase model enzymes. None of the isolated HOSPIs was active as inhibitor of subtilisin, while all HOSPIs inhibited the proteolytic activity of trypsin. However, the inhibitory constants differed as $K_i$ values between 0.22 and 167 nM (Table 1). HOSPI-C1 showed the strongest inhibitory activity ($K_i = 0.22$ nM), whereas HOSPI-B1 and B4, which contain the cleavage site on the first reactive loop and, exhibited the lowest trypsin inhibitory activity ($K_i = 167$ and 110 nM, respectively). HOSPIs isolated from DEAE-fraction A showed stronger inhibition than HOSPIs from DEAE-fraction B (Table 1, Figure 1A). All HOSPIs, except HOSPI-B1 showed inhibition of chymotrypsin, but the inhibitory potencies were about 10-fold weaker as compared to trypsin inhibition. The weakest activity was observed in HOSPI-B4 and -C1 ($K_i = 1.2$ mM), and HOSPI-B1 did not inhibit chymotrypsin. HOSPI-A6 and -A7 also inhibited elastase strongly, the $K_i$ value in case of HOSPI-A6 was 4.8 nM, while HOSPI-A7 could not determine it due to a low amount of native protein despite the activity similar trend with HOSPI-A6 (not shown in the table, we investigated details in a mutant analysis). Interestingly, HOSPI-A6 and -A7 which has Leu-Met at the putative reactive site exhibited the highest inhibitory activity against trypsin, $\alpha$-chymotrypsin and elastase. This is a unique sequence motif Leu-Met at the P1-P1' position in trypsin inhibitors (more details are described in the discussion section). The inhibitory activity profile of HOSPIs revealed that they have varying potencies against the different serine proteases.

**Cloning of cDNA precursor proteins encoding HOSPIs**

We isolated 15 novel serine proteinase inhibitors from hyacinth bulbs with different inhibitory activities. As an experimental approach to investigate the biosynthesis of these inhibitors in the plant, the gene structures of their precursors was elucidated. The cDNA was obtained via reverse transcription using an adaptor-linked oligo (dT) primer from the hyacinth bulb RNA. Six clones encoding HOSPI precursor proteins (HOSPI-1-1, -1-3, -1-8, -2-4, -2-6, and -x3) were analyzed by nucleotide sequencing, to determine the translated amino acid sequences (Figure 3A). All cDNA
clones except 

HOSPI-x3 contained two HOSPI-domains (HOSPI-ω 1 and -ω 2, Figure 3B) and an additional sequence not encoding for the mature inhibitor, which we referred to as linker region. In addition, these precursors contained a C-terminal elongation, which we referred to as C-tail region. The amino acid sequences between the mature inhibitor domains ω 1 and ω 2 were similar, but not identical. In addition, HOSPI-ω 1 in HOSPI-1-1 had high similarity with native HOSPI-A7 (98% identity).

The generated HOSPI-like cDNA clones were classified into two groups based on similarity of the sequence of mature domain ω1. ‘type X’ contained (i) a Leu-Met motif in the putative reactive site, (ii) 10 cysteine residues and (iii) a WPPVE motif at the N-terminus of the inhibitor coding sequence as observed in HOSPI-1-1, HOSPI-1-2, HOSPI-x3 (Figure 3A). Comparison of the linker regions between HOSPI-1-1 and HOSPI-1-2, with 21 or 24 residues, respectively, suggested a deletion of three amino acids, Asp-His-Ser (Figure S3). The C-tail region consists of 19 residues with high similarity (84% identity) among the three genes (Figure 3A).

A second group of precursor genes, termed ‘type Y’ contained (i) a Trp-Val in the putative reactive site, (ii) 12 cysteine residues and (iii) short type starting at Gly on N-terminus coding inhibitors as observed in HOSPI-1-3, HOSPI-1-8, and HOSPI-2-6. The linker regions in this group are much longer i.e. 45 residues with 95% identity. However, HOSPI-1-3 and HOSPI-1-8 do not contain a C-tail region (Figure 3A). The main finding of cloning the six HOSPI is that there are two distinct types of genes containing HOSPI-like domains in at least double repeats with a signature linker sequence connecting these domains.

Recombinant expression of HOSPI precursors in Escherichia coli

The HOSPI-ω1 gene products encoded in the C-terminal region of the two types (X or Y) of cDNA clones were investigated for their proteinase inhibitory activity. Specifically, it was of interest to investigate proteinase selectivity, and possible effects of the linker sequence. Therefore two representative clones, HOSPI-1-1 and HOSPI-1-8 (Figure 3A), were inserted into the expression vector pET32a and the fusion protein was produced using an E. coli expression system, (termed HOSPI-x1 and HOSPI-y1, Figure 4A). The sequences of the mature inhibitor HOSPI-X1 and Y1
encoded on LOSPI-x1 and LOSPI-y1 were compared with LOSPI-A7 of native inhibitor isolated from hyacinth bulb. As described above, HOSPI-A7 had a unique sequence motif, Leu-Met, at the P1-P1' position in the putative reactive site. The difference between HOSPI-A7 and HOSPI-X1 was Leu59 or Gin59 (Figure 4B). On the other hand, HOSPI-Y1 was confirmed on the level of nucleic acid, but the native inhibitor has not yet been identified. Remarkably, the putative reactive site of HOSPI-Y1 is Trp20-Val21, which would be unique in BBI. The HOSPI-X1 and -Y1 were expressed as a fusion with thioredoxin-tag, His-tag, and S-tag derived of pET32a vector. Hereinafter referred to as, a protein expressed with linker sequence (underlined in Figure 4A) is HOSPI-X1 (or Y1)-lk, while a protein without the linker sequence is HOSPI-X1-mat (or Y1-mat). The recombinant inhibitors were obtained in soluble form, and their size of approximately 32 kDa and 30 kDa was confirmed by SDS-PAGE analysis, respectively. The purified fusion proteins were obtained by affinity chromatography using a Ni2+-sepharose column and subsequent reverse-phase HPLC (Figure S4).

HOSPI-X1-lk and HOSPI-X1-mat showed strong trypsin inhibition within the same magnitude, $K_i$ values 7.0, or 7.7 nM, respectively (Figure 5A). HOSPI-X1 was also a strong inhibitor of chymotrypsin whether with or without linker sequence (Figure 5B), as shown by the $K_i$ values of 13 to 14 nM (Table 2). HOSPI-X1-lk and HOSPI-X1-mat showed a strong inhibition against elastase, as shown by the $K_i$ values of 6.3 or 3.1 nM, respectively (Figure 5C, Table 2). However, they showed no inhibition for subtilisin a known serine proteinase derived from Bacillus sp. (data not shown). These results showed similar extent of inhibitory activity to HOSPI-A7 (Table 1). In contrast, HOSPI-Y1-lk and -mat did not inhibit chymotrypsin or trypsin (Table 2). Thus, the linker sequences encoded on HOSPI-x1 and HOSPI-y1 were irrespective of activity against representative serine proteinases (Figure 4A).

**Sequence-activity studies of designed HOSPIs**

We designed mutant inhibitors based on HOSPI-X1-lk by varying amino acid residues around the putative reactive site. First, the unusual Leu20-Met21 at the P1-P1' position of HOSPI was interesting to mutate based on amino acids preference for trypsin, chymotrypsin or elastase inhibitions on the canonical BBI motif. Thus, the putative reactive site Leu20-Met21 was changed to Ala-Met, Leu-Ala,
Ala-Ala, and Arg-Ser by mutation experiments, as indicated with HOSPI-X1-AM, HOSPI-X1-LA, HOSPI-X1-RS in Table 2. Second, the residues of Arg\textsuperscript{51}-Ser\textsuperscript{52} located in the second canonical reactive site observed in dicot BBIs, Arg\textsuperscript{51} was changed to Ala, as named mutant HOSPI-X1-R51A. Finally, in an alanine scan of the residues next to the reactive site, Thr\textsuperscript{19}, Pro\textsuperscript{23} and Pro\textsuperscript{24}, were point mutated, as indicated with the probes HOSPI-T19A, HOSPI-X1-P23A, and HOSPI-X1-P24A in Table 2, respectively.

Regarding trypsin inhibitory activity, HOSPI-X1-LA, -RS, -AS showed the same level of inhibition as HOSPI-X1-1k. While HOSPI-AM activity was 3.7-times lower and -AA was 32.4-times lower than wild type (Table 2). In case of mutations in the vicinity of reactive site, in HOSPI-T19A, -P23A, and -P24A, their inhibitory activity were significantly reduced, 43-, 11-, 17-times, respectively. Regarding chymotrypsin inhibition, these changes of the reactive site and the vicinity residues revealed mild impact on the proteinase activity. The changes of the P1 site and Pro of 23- and 24-positions induced activities approximately three-times lower. In elastase inhibition, HOSPI-X1-RS, -T19A and -P23A showed no inhibitory activity, and HOSPI-AM and -AA slightly decreased activity. These results revealed that the residues on P1, P2, P3’ sites are pivotal interaction sites to maintain inhibition of elastase. All these mutants showed no inhibition for the subtilisin (data not shown).

We expected to recover part of the inhibitory activity of HOSPI-Y1 by designing an improved inhibitor using the putative reactive site sequence based on the Leu-Met motif in HOSPI-X1 and native HOSPI. The four mutants were prepared as follows; the mutant HOSPI-Y1-LM contained the reactive site of HOSPI-X1, the mutant HOSPI-Y1-RV had the Trp residue on the P1 site replaced to Arg, the mutant HOSPI-Y1-S19T had a Ser\textsuperscript{19} changed to Thr based on result of mutant analysis of LHTI-X1, and the mutant HOSPI-Y1-Y22A had the Tyr\textsuperscript{22} changed to Ala. However, they did not show inhibitory activity against trypsin and chymotrypsin, except HOSPI-Y1-S19T. Interestingly, the mutant HOSPI-Y1-S19T yielded inhibition against chymotrypsin with a $K\text{}_{\text{I}}$ value of 33 nM (Figure 5D). These results suggested that the linker sequence does not affect the inhibitory activity and a Thr residue at the P2 site as well as a Pro at the P3’ site are crucial for elastase inhibition.
Discussion

Among the proteinase inhibitors derived from plants, BBI constitute a large and diverse family [26]. BBIs are serine proteinase inhibitors abundantly expressed in dicotyledonous plants. BBIs from dicotyledonous plant have two reactive sites (double-headed) and a molecular weight of about 8 kDa [27]. On the other hand, BBIs from monocotyledonous plants can be classified into two groups: one that has a single reactive site and a molecular weight of < 8 kDa (they lack the second reactive site), and the other group that has two reactive sites with a molecular weight of < 16 kDa (i.e. two 8 kDa repeats) [28]. The reactive site is located in a unique loop connected by two β-sheets and a disulfide bridge [29]. This loop structure is a common feature in BBI family proteins and contributes to stabilization of the enzyme inhibition motif (Figure S1).

In this study, 15 HOSPIs from hyacinth bulbs with a molecular weight between 6.5 - 8.0 kDa were isolated. By homology to known inhibitors they have been classified as BBIs with conserved cysteine residues. In HOSPIs, the second reactive site conserved in other dicot plants, including soybean BBI, was not present due to a missing disulfide bond in the inhibitory loop (Figure 2). Remarkably, HOSPI-B1 and HOSPI-B4 have a six-residue reactive loop with cleavage site in the first reactive site loop similar to "modified inhibitors" [30]. This is the smallest inhibitor motif in BBIs that have been reported so far, and to the best of our knowledge the first of its kind observation in monocot BBIs.

Previously, we isolated a BBI from onion. The inhibitory activity of the onion BBI against chymotrypsin was 100-times weaker than the inhibition of trypsin. We demonstrated that Arg17 in the onion BBI is an important residue for inhibitory activity against trypsin, and predicted that Leu46 might be the responsible P1 site residue for chymotrypsin inhibition even though the onion BBI lacks a disulfide bond as observed in double-headed BBIs [16]. In this study, seven of nine LOSPIs for which we determined the complete amino acid sequence have an Arg in the P1 site of the first reactive loop, and they exhibit varying degrees of inhibitory activity, i.e. the $K_i$ values for trypsin inhibition were 4 to 5,000 -times higher than for chymotrypsin inhibition (Table 1). This suggested that the reactive site in native LOSPIs for trypsin inhibition is the Arg in the first reactive loop. Although the reactive site for chymotrypsin hasn’t been determined by a comparison with these native inhibitors,
HOSPIs with the Thr at P2 site and Leu at P1 site strongly inhibited trypsin and chymotrypsin (Table 1).

We found unique inhibitors, HOSPI-A6 and -A7 with the reactive site of Leu-Met (P1-P1'), showed strong trypsin, chymotrypsin, and elastase inhibitions, a further unprecedented observation from our study (Table 1). In well studied double-headed BBIs from dicot, the P1 amino acid residue of the first reactive site is Lys or Arg for trypsin inhibition, whereas the second one is Arg, Phe, Tyr, Leu or Ala for trypsin, chymotrypsin or elastase inhibitions. Additionally, the P1' residue in the first and the second site is a conserved Ser. According to analysis of inhibitor derivatives, the preference for chymotrypsin inhibition decreases with the P1 site residues Phe > Trp > Leu > Met > Val > Ala > Gly [31], while P1' site residues are Ser > Ala > Thr > Val > Leu > Gly [32]. To date, no native BBIs with Leu-Met at the P1-P1' site have been reported, thus the herein reported HOSPIs may have a unique reactive site sequence. Therefore, we tried to examine the reactive site residues of the HOSPIs by preparing recombinant proteins for mutant analysis.

In the case of BBIs, the importance of the P3' residue is known; Pro at the P3' site of BBIs isolated from legumes is conserved. This Pro residue adopts a cis peptide bond and as a result the reactive site loop forms a type VIb β-turn [29, 33]. The inhibitory activity is significantly reduced if the Pro residue is replaced by other amino acids. In contrast, the P3' sites of LOSPIs are not conserved; all HOSPI B-proteins (except HOSPI-B3) and -C1 are substituted with other amino acids in this position. The lack of the conserved proline might be an explanation for the low inhibitory activity of the HOSPI-Bs.

During analysis of cDNAs encoding HOSPI precursors, we observed many duplications and tandem repeat sequences. Although it is notably cumbersome due to similarity of nucleotide sequences, eventually we determined a partial nucleotide sequence of six clones, one of which encodes the HOSPI-X has 10 cysteine residues, and the other encodes HOSPI-Y has 12 cysteine residues in the C-terminal domain. HOSPI-X1 has only Leu\textsuperscript{60} of native HOSPI-A7 replaced by Glu\textsuperscript{60}, yet the determined amino acid sequences of these cDNA clones have not been found in the native mature inhibitor. Three remarkable points of these HOSPI gene sequence are: (1) the additional linker and tail regions have not been found in native proteins, (2) like native HOSPI sequences are repeated
at least twice, and (3) an unusual sequence of a putative reactive site (WV) was found in the *HOSPI-Y1* gene. Accordingly, *HOSPI-X1-lk* and *HOSPI-X1-mat* were prepared as fusion proteins to evaluate the contribution of the linker sequences to activity. The *HOSPI-X1-lk* inhibited three serine proteinases similar as did the native type. Hence, it can be concluded that the linker sequence does not affect the inhibitory activity of *HOSPI-X1*. On the other hand, *HOSPI-Y1* did not exhibit serine proteinase inhibitory activity irrespective of the presence or absence of the linker sequence, which might imply that the linker is not crucial for serine proteinase inhibition. A Trp in the P1 site found in *HOSPI-Y1* has not yet been reported for native BBIs. In mutational studies of the smallest known BBI peptide, the cyclic sunflower trypsin inhibitor (SFTI-I, cyclo-GRCTKSIPPICFPD), the Lys to Trp mutant in the P1 site shifts the inhibitory activity towards chymotrypsin [34, 35]. Our results suggested that the Thr at the P2 site contributes to chymotrypsin inhibition, in agreement with the SFTI-I study.

The analysis of *HOSPI-X1* further revealed the broad specificity against trypsin, elastase, and chymotrypsin. The inhibitory activity for trypsin and elastase were significantly decreased by mutations around the reactive site, especially the P2 site of *HOSPI-X1*. Additionally, the mutant P23A (Pro residue at the P3’ site was replaced by Ala) showed low activity. These results suggested that the surrounding residues of the reactive site might contribute to the recognition and/or interaction with the enzymes. For instance, the type VIb β-turn supports inhibition activity [29, 33]. Thus, the proteinase inhibitory activity of *HOSPI* is not only dictated by the few amino acids of the reactive site, but also by a larger segment usually referred to as binding face in BBI, which are surrounding amino acids of the reactive site or neighboring loops.

Regarding the observed *HOSPI* gene structure, there is only one example of a cysteine proteinase inhibitor, bromelain inhibitor from pineapple stem, in which multiple inhibitor domains are connected in tandem in the precursor [36]. In this case, three isoinhibitors are encoded in the gene in series connected by a linker sequence between the domains. It has been hypothesized that these mature forms of bromelain isoinhibitors may be processed because of the cooperativity of certain proteinases, leading to a variation in the processing sites [37]. The amino acid sequences of the linker in the *HOSPI* genes are different to the genes of bromelain isoinhibitors. Supposedly, at least two
processing enzymes having different substrate specificity will be necessary to process the precursor peptide into the mature HOSPI. Since the boundary of the HOSPI-X1 and the tail sequence is almost the same amino acid sequence as the N-terminal on the linker sequence of HOSPI-X1, the tail sequence might be separated by the same processing enzyme. We have not yet isolated the enzyme that processes the HOSPI precursor, however, our preliminary data (not shown) suggested that this enzyme belongs to the aspartic acid proteinases.

In summary, we isolated 15 novel BBIs from the bulb of *H. orientalis*, which potently inhibited mammalian serine proteinases. Notably, one of them contains the smallest inhibitory motif in BBIs discovered to date, and the others contain an unusual reactive site at the P1-P1’ position resulting in distinct substrate preference. Studying recombinant inhibitor mutants led to the conclusion that the key residues for the observed the activity profile were Pro residues at the P2’ and P3’, respectively, and a Thr residue at the P2 site. Our data suggest that HOSPIs might interact with nearby residues of the catalytic site to build a shield in the enzyme’s pockets. An ongoing three-dimensional structure analysis of a HOSPI in complex with a mammalian serine proteinase will clarify more molecular details of the inhibitor/enzyme recognition and inhibition mechanism. This might advance the design of selective inhibitors in the future.

At a more general level, BBI proteins are considered for their applications in drug development and agriculture. Firstly, they are emerging drug candidates for development of novel therapeutics of cancer and inflammatory diseases [6]. Secondly, it is known that BBI gene expression is activated as a response of the plants’ innate immune system following herbivore attack or microbial infection. Therefore, BBIs are important plant defense molecules that may also be engineered in transgenic plants [38]. In this study, it was possible to recombinantly express functional HOSPI inhibitors in *E. coli* that target several different proteinases. This is a first step to demonstrate feasibility for upscaling production of these novel BBIs. By modifying the non-active HOSPI-Y1 it was possible engineer a desired functionality. This can be considered the groundwork to design more valuable HOSPI-based protein inhibitors in the future to be used, for example, as insecticides or therapeutics.
Acknowledgments

We thank Dr. Masanobu Deshimal for providing professional advices about HOSPI gene analysis of plant. We also thank Mrs. Rie Maenaka and Mrs. Yuki Uemura for helpful sample preparations. We appreciate staffs of Radioisotope Centre of Fukuoka University for the analysis of protein sequences in this study.

Funding and additional information

This work was supported by Discovery Grants by JSPS KAKENHI Grant Number 17KK0179 and JBC31025 from Japanese government. This research was supported by grant P32109 of the Austrian Science Fund (FWF). This work was supported by Central Research Institute of Fukuoka University (181042 and 197104) supported by the female researcher support programme of Fukuoka University to N.S.

Conflict of Interest: The authors have no conflicting financial interests.

Data Availability Statement: Protein sequences of the Hyacinthus orientalis serine proteinase inhibitors have been submitted to the UniProt Knowledgebase as follows; C0HLS6 (HOSPI-A4), C0HLS7 (HOSPI-A6), C0HLS8 (HOSPI-A7), C0HLS9 (HOSPI-B1), C0HLT0 (HOSPI-B4), C0HLT1 (HOSPI-B5), C0HLT2 (HOSPI-B6), C0HLT3 (HOSPI-B7), and C0HLT4 (HOSPI-C1).

The abbreviations used are: BBI, Bowman-Birk family inhibitor; Boc, t-butyloxycarbonyl; Bz, benzoyl; MALDI-TOF-MS, matrix-assisted laser desorption-ionization time-of-flight mass spectrometry; MCA, 4-methylcoumaryl-7-amide; HOSPI, Hyacinthus orientalis serine proteinase inhibitors; Pe, S-pyridylethylated; pNA, p-nitroanilide; Suc, succinyl
References

1. Laskowski, M., Jr. and Kato, I. (1980) Protein inhibitors of proteinases. Annu Rev Biochem. 49, 593-626
2. Hellinger, R. and Gruber, C. W. (2019) Peptide-based protease inhibitors from plants. Drug Discov Today. 24, 1877-1889
3. Kunitz, M. (1947) Isolation of a Crystalline Protein Compound of Trypsin and of Soybean Trypsin-Inhibitor. J Gen Physiol. 30, 311-320
4. Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A. and Finn, R. D. (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 46, 624-632
5. Kennedy, A. R. (1998) The Bowman-Birk inhibitor from soybeans as an anticarcinogenic agent. Am J Clin Nutr. 68, 1406-1412
6. Clemente, A. and Arques Mdel, C. (2014) Bowman-Birk inhibitors from legumes as colorectal chemopreventive agents. World J Gastroenterol. 20, 10305-10315
7. Scarpi, D., McBride, J. D. and Leatherbarrow, R. J. (2004) Inhibition of human beta-trypsin by Bowman-Birk inhibitor derived peptides: creation of a new tri-functional inhibitor. Bioorg Med Chem. 12, 6045-6052
8. Odani, S., Odani, S., Ono, T. and Ikenaka, T. (1979) Proteinase inhibitors from a mimosoideae legume, Albizia julibrissin. Homologues of soybean trypsin inhibitor (Kunitz). J Biochem. 86, 1795-1805
9. Birk, Y. (1985) The Bowman-Birk inhibitor. Trypsin- and chymotrypsin-inhibitor from soybeans. Int J Pept Protein Res. 25, 113-131
10. Bowman, D. E. (1944) Fractions derived from soy beans and navy beans which retard tryptic digestion of casein. P Soc Exp Biol Med. 57, 139-140
11. Ryan, C. A. (1990) Protease Inhibitors in Plants - Genes for Improving Defenses against Insects and Pathogens. Annu Rev Phytopathol. 28, 425-449
12. Prakash, B., Selvaraj, S., Murthy, M. R., Srerama, Y. N., Rao, D. R. and Gowda, L. R. (1996) Analysis of the amino acid sequences of plant Bowman-Birk inhibitors. J Mol Evol. 42, 560-569
13. Park, E. Y., Kim, J. A., Kim, H. W., Kim, Y. S. and Song, H. K. (2004) Crystal structure of the Bowman-Birk inhibitor from barley seeds in ternary complex with porcine trypsin. J Mol Biol. 343, 173-186
14. Qasim, M. A., Ganz, P. J., Saunders, C. W., Bateman, K. S., James, M. N. and Laskowski, M., Jr. (1997) Interscaffolding additivity. Association of P1 variants of eglin c and of turkey ovomucoid third domain with serine proteinases. Biochemistry. 36, 1598-1607
15. Lu, W., Apostol, I., Qasim, M. A., Warne, N., Wynn, R., Zhang, W. L., Anderson, S., Chiang, Y. W., Ogin, E., Rothberg, I., Ryan, K. and Laskowski, M., Jr. (1997) Binding of amino acid side-chains to S1 cavities of serine proteinases. J Mol Biol. 266, 441-461
16. Deshimaru, M., Watanabe, A., Suematsu, K., Hatano, M. and Terada, S. (2003) Purification, amino acid sequence, and cDNA cloning of trypsin inhibitors from onion (Allium cepa L.) bulbs. Biosci Biotech Bioch. 67, 1653-1659
17. Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem. 182, 319-326
Morrison, J. F. (1969) Kinetics of the reversible inhibition of enzyme-catalysed reactions
by tight-binding inhibitors. Biochim Biophys Acta. 185, 269-286
Copeland, R. A. (2013) Evaluation of Enzyme Inhibitors in Drug Discovery. John Wiley
& Sons, Inc.
Williams, J. W. and Morrison, J. F. (1979) The kinetics of reversible tight-binding
inhibition. Methods Enzymol. 63, 437-467
Lineweaver, H. and Burk, D. (1934) The determination of enzyme dissociation constants.
J Am Chem Soc. 56, 658-666
Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of
bacteriophage T4. Nature. 227, 680-685
Friedman, M., Krull, L. H. and Cavins, J. F. (1970) The chromatographic determination
of cystine and cysteine residues in proteins as s-beta-(4-pyridylethyl)cysteine. J Biol Chem. 245,
3868-3871
Falco, M. C. and Silva-Filho, M. C. (2003) Expression of soybean proteinase inhibitors in
transgenic sugarcane plants: effects on natural defense against Diatraea saccharalis. Plant Physiol
Bioch. 41, 761-766
Birk, Y., Gertzler, A. and Khalef, S. (1963) A pure trypsin inhibitor from soya beans.
Biochem J. 87, 281-284
Rawlings, N. D., Tolle, D. P. and Barrett, A. J. (2004) Evolutionary families of peptidase
inhibitors. Biochem J. 378, 705-716
Odani, S. and Ikenaka, T. (1973) Scission of soybean Bowman-Birk proteinase inhibitor
into two small fragments having either trypsin or chymotrypsin inhibitory activity. J Biochem.
74, 857-860
Qi, R. F., Song, Z. W. and Chi, C. W. (2005) Structural features and molecular evolution
of Bowman-Birk protease inhibitors and their potential application. Acta Biochim Biophys Sin
(Shanghai). 37, 283-292
Brauer, A. B., Kelly, G., McBride, J. D., Cooke, R. M., Matthews, S. J. and
Leatherbarrow, R. J. (2001) The Bowman-Birk inhibitor reactive site loop sequence represents an
independent structural beta-hairpin motif. J Mol Biol. 306, 799-807
Niekamp, C. W., Hixson, H. F., Jr. and Laskowski, M., Jr. (1969) Peptide-bond
hydrolysis equilibria in native proteins. Conversion of virgin into modified soybean trypsin
inhibitor. Biochemistry. 8, 16-22
Odani, S. and Ono, T. (1980) Chemical substitutions of the reactive site leucine residue in
soybean Bowman-Birk proteinase inhibitor with other amino acids. J Biochem. 88, 1555-1558
Odani, S. and Ikenaka, T. (1978) Studies on soybean trypsin inhibitors, XII. Linear
sequences of two soybean double-headed trypsin inhibitors, D-II and E-I. J Biochem. 83, 737-745
Brauer, A. B., Domingo, G. J., Cooke, R. M., Matthews, S. J. and Leatherbarrow, R. J.
(2002) A conserved cis peptide bond is necessary for the activity of Bowman-Birk inhibitor
protein. Biochemistry. 41, 10608-10615
McBride, J. D., Freeman, H. N. and Leatherbarrow, R. J. (2000) Identification of
chymotrypsin inhibitors from a second-generation template assisted combinatorial peptide
library. J Pept Sci. 6, 446-452
McBride, J. D., Watson, E. M., Brauer, A. B., Jaulent, A. M. and Leatherbarrow, R. J.
(2002) Peptide mimics of the Bowman-Birk inhibitor reactive site loop. Biopolymers. 66, 79-92
36 Sawano, Y., Hatano, K. and Tanokura, M. (2005) Susceptibility of the interchain peptide of a bromelain inhibitor precursor to the target proteases bromelain, chymotrypsin, and trypsin. Biol Chem. 386, 491-498

37 Sawano, Y., Muramatsu, T., Hatano, K., Nagata, K. and Tanokura, M. (2002) Characterization of genomic sequence coding for bromelain inhibitors in pineapple and expression of its recombinant isoform. J Biol Chem. 277, 28222-28227

38 Kennedy, A. R. (1998) Chemopreventive agents: protease inhibitors. Pharmacol Ther. 78, 167-209
## Tables

### Table 1. Inhibitory activity ($K_i$) of isolated native HOSPIs against trypsin and chymotrypsin.

| Inhibitors      | Residues in vicinity of the reactive site$^a$ | $K_i$ (nM) | trypsin | chymotrypsin |
|-----------------|---------------------------------------------|-----------|---------|--------------|
|                 | P_1 P_2 P_3 P_4 P_5 P_6 P_7 P_8 P_9 |           |         |              |
| HOSPI-A4        | ACTRMPPP                                           | 12        | 460     |              |
| HOSPI-A6        | ACTLMPD                                              | 4.0       | 22      |              |
| HOSPI-A7        | ACTLMPPP                                             | 7.1       | 19      |              |
| HOSPI-B1        | ACARSLCI                                             | 167       | -b      |              |
| HOSPI-B4        | ACARSLCI                                             | 110       | 1200    |              |
| HOSPI-B5        | ICDRSLDP                                             | 41        | 410     |              |
| HOSPI-B6        | ICDRSLDP                                             | 30        | 550     |              |
| HOSPI-B7        | ICGRAGFS                                             | 65        | 295     |              |
| HOSPI-C1        | ICDRSGDP                                             | 0.22      | 1200    |              |

$^a$, putative reactive residue based on a canonical BBI domain as shown in Supplementary Figure S1; $^b$, inactive
Table 2. Inhibitory constants ($K_i$) of HOSPI-X1 and -Y1 expressed in *E. coli*.

| HOSPIs     | Residues in vicinity of the reactive site | $K_i$ (nM) | trypsin | chymotrypsin | elastase |
|------------|------------------------------------------|------------|---------|-------------|----------|
|            | $P_3P_2P_1P_1'$ |            |         |             |          |
| HOSPI-X1-lk| CTLMWPPP          | 7.0        | 13      | 6.3         |          |
| HOSPI-X1-mat| CTLMWPPP         | 7.7        | 14      | 3.1         |          |
| HOSPI-X1-AM| CTAMWWPP          | 26         | 43      | 17          |          |
| HOSPI-X1-LA| CTLMWWPP          | 7.0        | 14      | 4.4         |          |
| HOSPI-X1-AA| CTAMWWPP          | 230        | 13      | 14          |          |
| HOSPI-X1-RS| CTRSWWPP          | 8.8        | 22      |             |          |
| HOSPI-X1-R51A | CTLMWWPP......$A_2S_2$ | 8.8 | 12 | 4.3 |          |
| HOSPI-X1-T19A| CAWLMPPP        | 300        | 17      |             |          |
| HOSPI-X1-P23A| CTLMWAP        | 77         | 39      |             |          |
| HOSPI-X1-P24A| CTLMWPPA        | 120        | 33      | 5.1         |          |
| HOSPI-Y1lk | CSWVYPF          | -          | -       | n.d.        |          |
| HOSPI-Y1-mat| CSWVYPF         | -          | -       | n.d.        |          |
| HOSPI-Y1-LM | CSTWVFP          | -          | -       | n.d.        |          |
| HOSPI-Y1-RV | CSTWVFP          | -          | -       | n.d.        |          |
| HOSPI-Y1-S19T| CSTWVFP         | -          | 33      | n.d.        |          |
| HOSPI-Y1-Y22A| CSGVFP          | -          | -       | n.d.        |          |

*a*, mutation residues are shown in bold; *b*, second reactive site observed in dicot BBI as shown in Supplementary Figure S1; *c*, (-) inactive; *d*, n.d., not determined. Residues are numbering based on Figure 4B.
**Figure 1. Purification of hyacinth trypsin inhibitors.** (A) Hyacinth bulb extract by anion exchange chromatography (DEAE-Toyoperal) was monitored by $A_{280}$ as shown in a solid line. Trypsin inhibitory activity of these fractions was indicated by a dashed line. (B - D) Reversed-phase HPLC of DEAE fractions A, B and C were carried out using a linear gradient of acetonitrile in 0.1% TFA from 0.1% TFA (dashed line) and monitored via $A_{230}$ (solid line).
**Figure 2. Amino acid sequences of hyacinth trypsin inhibitors (HOSPIs).** An alignment of novel HOSPI sequences (A1-A7, B1-B7, and C1) is presented in comparison to onion, wheat, and soybean BBIs (bottom), their accession numbers are BAB88746.1, AAB31685.1 and A0A0B2NX54, respectively. Cysteine residues are highlighted in dark, and the first and second reactive site loops (inhibitory loops) based on soybean BBI are indicated by a connecting line (see Figure S1). The reactive sites (P1-P1’ sites) against serine proteinases are marked with asterisks. Sequence gaps are indicated by hyphens. Black triangular marks indicate the proposed cleavage site of the separate chains linked by disulfide bonds. The N-terminal region of HOSPI-A2, -A3, -A5 and -B3 wasn’t determined completely, therefore “x” indicates a partial sequence. The N-terminal residue of HOSPI-A7 is pyroglutamic acid and has been indicated with a superscript of “py”. These complete or a partial amino acid sequences of native HOSPIs were assembled using the peptide sequences determined by Edman sequencing of peptides obtained by digestion of these proteins by several endoproteinases (see also Figure S2).

| **HOSPI-A2** | **HOSPI-A3** | **HOSPI-A4** | **HOSPI-A5** | **HOSPI-A6** | **HOSPI-A7** | **HOSPI-B1** | **HOSPI-B3** | **HOSPI-B4** | **HOSPI-B5** | **HOSPI-B6** | **HOSPI-B7** | **HOSPI-C1** | **Onion BBI** | **Wheat BBI** | **Soybean BBI** |
|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|----------------|
| TWPVGERPSGRKGDQACRxx | TWPVGERPSGRKGDQACRxx | WPPVEEPGRPSGRKGDQACRxx | WPPVEEPGRPSGRKGDQACRxx | WPPVEEPGRPSGRKGDQACRxx | WPPVEEPGRPSGRKGDQACRxx | QQPQLNQGAFACRSLG---GICGDLPQ-GCHGDCEKVDLSQPKLYQGRSFEDYGxx | GQQPLNQGAFACRSLG---GICGDLPQ-GCHGDCEKVDLSQPKLYQGRSFEDYGxx | GQQPLNQGAFACRSLG---GICGDLPQ-GCHGDCEKVDLSQPKLYQGRSFEDYGxx | GRRPCNQEGCICORSLDP---NGCDELPQ-GCHGDCEEVDTQTPMYQGRSFEEYHGTPQCL | GRRPCNQEGCICORSLDP---NGCDELPQ-GCHGDCEKVDTSQPKLYQGRSFEDYGxx | GRRPCNQEGCICORSLDP---NGCDELPQ-GCHGDCEKVDTSQPKLYQGRSFEDYGxx | E-RPLQNEF-FCQSGDP---RLCQEDHVPQ-GCHGDCEKVDTSQPKLYQGRSFEDYGxx | GDEEEGCGCSSGDSOARPAVELRCEQERDVTSSQCCRPCRCEADLDQPRYYXOMSFHSQOTTROSIL | EEAPGSAWPCRECG-GTCGTRMIPPP---RTGDVVSPSOGGPAVRQUTTLGQRVFDW---MLR1ENFODRKRPAR | DDESSKPCOCQ---AATGSNPP---QCGCDRLNSQHACKSNPSIALSYPP---FQVDT-DFYEPKXPSDDKEN |

**First reactive site loop**
**Second reactive site loop**
(for soybean BBI only)
Figure 3. Sequencing of HOSPI precursors. (A) The amino acid sequence determined from the nucleotide sequence of the six cDNA clones encoding HOSPI. The primer (5'-TTGGTTCCACAATGCCACGAGGCTGTC-3') used for PCR aligns to N-terminal part of sequence (Val-Pro-Gln-Cys-His-Glu-Gly-Cys) and has been highlighted in boldface. The regions corresponding to native HOSPI are marked in grey. The underlined parts belong to the linker sequence. The dotted underlined parts indicate the tail sequence. Sequence gaps of three amino acid residues exist in comparison between the linker sequences of HOSPI-1-1 and HOSPI-2-6, and they are indicated by hyphens. Downstream sequences including stop codons in these cDNA clones were determined, but start codons were not detected in these cDNA clones. Thus, the partial sequences are indicated with “xxx”. (B) Schematic diagram of precursors of HOSPI obtained by HOSPI gene sequence of cDNA clones. The provisional HOSPI-like region name was ω1 and ω2 in direction from the C- to N-terminus.
Figure 4. Recombinant HOSPI analogues. (A) Nucleotide-based design of recombinant HOSPIs using the pET 32a vector. Parts of the amino acid sequence from the cDNA clones of HOSPI-x1 and HOSPI-y1 were fused with tags, a thioredoxin protein (Trx-tag), His-tag, and S-tag, derived from the expression vector. Uppercase letters are corresponding to HOSPI (BBI domain) and the lower-case underlined sequences indicate the linker. In case of HOSPI-x1, the mature form has high sequence similarity with native HOSPIs, and the precursor form including a linker sequence was prepared. In case of HOSPI-y1, the cloned sequence has not been found in the native protein. Thus, the expression vectors of the mature form of HOSPI-Y1 was determined from sequence re-arrangement based on the conserved cysteine residues. The mature HOSPI-Y1 was prepared as a short-type of N-terminal HOSPIs by similarity of native protein in Figure 1. (B) The comparative sequences of mature HOSPI-X1, -Y1 and native HOSPI-A7 have been aligned. Numbering of amino acid sequence in this alignment was based on the native HOSPI-A7. The putative reactive site residues (P1, P1’) on the inhibitor loop are shown in asterisks. The underlined sequences of R51-S52 are showed a second reactive site observed in dicot BBI (see Figure S1).
Figure 5. Trypsin, chymotrypsin, and elastase inhibitory activities of recombinant by HOSPI-X1s.

(A) Inhibitory activities of HOSPI-X1 with (HOSPI-X1-lk) and without a linker (mature form; HOSPI-X1-mat) were assayed against trypsin using 0.08 μM trypsin, 0.25 mM fluorogenic substrate (L-BAPA), and varying concentrations of inhibitor. (B) Inhibitory activities of HOSPI-X1s against α-chymotrypsin using 0.45 μM chymotrypsin, 0.15 mM fluorogenic substrate (Bz-Tyr-pNA), and varying concentrations of inhibitor. (C) Inhibitory activities of HOSPI-X1s against elastase using 0.05 μM elastase, 0.25 mM fluorogenic substrate (SAPA), and varying concentrations of inhibitor. Concentration-dependent inhibition of HOSPI-X1-lk and HOSPI-X1-mat was shown with closed and open circles, respectively. (D) Inhibitory activity of HOSPI-Y1-S19T against α-chymotrypsin using 0.45 μM chymotrypsin, 0.15 mM fluorogenic substrate (Bz-Tyr-pNA), and varying concentrations of inhibitor. In all figures, residual activity at the y-axis is corresponding to the fractional activities ($V_i/V_0$) at each inhibitor concentration, which were calculated by using the velocity of the uninhibited reaction ($V_0$, without inhibitor). The value of $K_{i}^{app}$ was obtained by non-linear regression of the data based on Morrison’s equation for tight binding inhibitors [31] (see Materials and Methods).
Supplementary Material for

Isolation and functional diversity of Bowman-Birk type serine proteinase inhibitors from *Hyacinthus orientalis*

Narumi Shioi†,*,1, Shigeyuki Terada†,1, Roland Hellinger2, Yoshitaka Furuta1, Christian W. Gruber2

1Department of Chemistry, Faculty of Science, Fukuoka University, 19-1, 8-chome Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan
2Center for Pharmacology and Physiology, Medical University of Vienna, Schwarzspanierstr. 17, 1090 Vienna, Austria

†These authors contributed equally to this work.

*To whom correspondence should be addressed: Narumi Shioi: Department of Chemistry, Faculty of Science, Fukuoka University, 19-1, 8-chome Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan; anarumi@fukuoka-u.ac.jp; Tel. +81-(0)92-870-6631 ext. 6215; Fax.+81-(0)92-865-6030;
Table of Contents

S 1. Title and table of contents
S 3. Supplementary tables S1 to S2
S 5. Supplementary figures S1 to S4
S10 References of the supplementary file.
Table S1. The average theoretical molecular weight (MW, molar extinction coefficient (ε M⁻¹cm⁻¹ at 280 nm), and isoelectric point (pI) of purified each HOSPI was calculated computationally based on the amino acid sequence as shown in Figure 2 by the ExPASy server (https://web.expasy.org/protparam/). In this list, assuming all pairs of cysteine residues form disulfide bonds.

| HOSPIs | The number of residues | Cysteine residues | MW (theoretical value, Da) | ε M⁻¹cm⁻¹ | pI | The number of chains | MW (Da) |
|--------|------------------------|-------------------|---------------------------|-----------|----|----------------------|---------|
| A4     | 67                     | 10                | 7579.66                   | 16095     | 7.64 | 1                    | 7677.2  |
| A6     | 67                     | 12                | 7444.53                   | 10720     | 4.84 | 1                    | 7488.9  |
| A7     | 70                     | 10                | 7957.19                   | 16095     | 8.10 | 1                    | 7997.2  |
| B1     | 55                     | 10                | 6110.91                   | 3605      | 4.65 | 3                    | 6142.1  |
| B4     | 57                     | 10                | 6311.15                   | 3605      | 4.65 | 2                    | 6322.8  |
| B5     | 61                     | 12                | 6858.80                   | 5220      | 4.48 | 2                    | 6859.6  |
| B6     | 61                     | 12                | 6919.93                   | 6710      | 4.82 | 2                    | 7012.3  |
| B7     | 62                     | 12                | 6895.82                   | 9230      | 4.51 | 1                    | 7240.1  |
| C1     | 60                     | 10                | 7002.74                   | 5095      | 4.61 | 1                    | 7248.0  |
**Table S2.** PCR primers designated based on amino acid sequences of the purified native inhibitors.

| Primer          | 5’ to 3’                                                      |
|-----------------|--------------------------------------------------------------|
| HOSPI5-N a)     | TGTYGAAYGARTGYGGNATHTGYGAY                                  |
| HOSPI6-N a)     | NTGYGAAYGARTGYGGNATHTGYGAY                                  |
| 3’-Adp b)       | GGCCACGCGTCTAGACTAGTAC                                       |
| Ex-IkN          | TATAACTGGCGCATGGACGGCACATCGT                                 |
| Ex-matN         | TTCAGAAGTGCGACTACATGGCGTGG                                  |
| Ex-matC         | TGGAAGCTTCCCTAAAGCGCGCAGTTATA                                 |

a) Mixed bases were indicated with underline by Y(C+T), R (A+G), N (N+C+G+T), and H (A+C+T).

b) Total RNA was reverse-transcribed into cDNA using an adaptor-linked oligo (dT) primer.
(A) Dicotyledonous plant (double-headed BBI)
Canonical BBI domain (6–8 kDa)

Canonical inhibitor motif (9–11 residues)

\[
\text{DDESSKPCDQCACAKSNNPQCRCSMDRLNSCHSACKSCCALSYPAQCFCVDITDFECYPCKPSEDDKEN}
\]

Monocotyledonous plant
Class I (~8 kDa, single reactive site)

\[
\begin{array}{ccccccc}
\text{C} & \text{C} & \text{C} & \text{X} & \text{C} & \text{C} & \text{C} \\
1 & 2 & 3 & 4 & 5 & 6 & 7
\end{array}
\]

Class II (~16 kDa, two BBI domains)

\[
\begin{array}{ccccccc}
\text{C} & \text{C} & \text{C} & \text{X} & \text{C} & \text{C} & \text{C} \\
1 & 2 & 3 & 4 & 5 & 6 & 7
\end{array}
\]

(B) Reactive loop

\[
\begin{array}{ccccccc}
\text{C} & \text{C} & \text{C} & \text{X} & \text{C} & \text{C} & \text{C} \\
1 & 2 & 3 & 4 & 5 & 6 & 7
\end{array}
\]

(C) Polypeptide chain

\[
\begin{array}{ccccccc}
\text{1C2A} & \\
10 & 4 & 5 & 6 & 7 & 8 & 9
\end{array}
\]

S5
**Figure S1.** Structural features of Bowman-Birk proteinase inhibitors (BBI). (A) The canonical BBI domain structure from dicotyldoneaous plants is shown based on soybean BBI (Glycine max, sequence databases ID; A0A0B2NX54) as a representative double-headed serine proteinase inhibitor. Highly conserved cysteine residues and their disulfide patterns in this domain are shown. The reactive sites (P1-P1’) against serine proteinases are indicated with asterisks and exist in inhibition loops (also termed reactive loop) consisted of disulfide bonds between C4-C5 and C10-C11. The schematic diagram of the BBI domain organisation isolated from monocotyledonous plants (exemplified by wheat BBI, Triticum aestivum, sequence databases ID; AAB31685) is shown in the lower panel. Cysteine residues C10-C11 in double-headed BBI from a disulfide bond (A), which were replaced by non-cysteine residues in monocots (indicated by “X”). In monocot BBIs, two different forms (class I and II) were identified and one reactive loop has been lost as indicated by the dotted line [1, 2]. (B) Crystal structure of soybean BBI (Glycine max, Protein data bank ID; 5J4Q). Cartoon representation of the double-headed BBI is shown. Two reactive sites are surface exposed by two inhibition loops highlighted in pink area of the tertiary structure. (C) Amino-acid sequence alignments and highly conserved disulfide pairings of BBIs. These proteins are listed on Protein data bank (PDB) ID and these number of amino acid resides are indicated in light panel. All conserved cysteine residues are shown in red. Disulfide pairings are indicated below the alignment with red numbers and these intramolecular disulfide linkages are shown in solid line. Their highly conserved disulfide patterns were observed by determined three-dimensional structures on PDB as follows: 1C2A (Cys6-Cys60, Cys7-Cys22, Cys12-Cys20, Cys29-Cys36, Cys33-Cys48, Cys65-Cys119, Cys66-Cys81, Cys71-Cys79, Cys88-Cys95, Cys92-Cys107), 5J4Q (Cys8-Cys62, Cys9-Cys24, Cys12-Cys58, Cys14-Cys22, Cys32-Cys39, Cys36-Cys51, Cys41-Cys49), 5J4S (Cys8-Cys62, Cys9-Cys24, Cys12-Cys58, Cys14-Cys22, Cys32-Cys39, Cys36-Cys51, Cys41-Cys49), 1PBI (Cys8-Cys61, Cys9-Cys24, Cys12-Cys57, Cys14-Cys22, Cys32-Cys38, Cys35-Cys50, Cys40-Cys48), 2AIH (Cys8-Cys61, Cys9-Cys24, Cys12-Cys57, Cys14-Cys22, Cys31-Cys38, Cys35-Cys50, Cys40-Cys48), 1PI2 (Cys7-Cys60, Cys8-Cys23, Cys11-Cys56, Cys13-Cys21, Cys30-Cys37, Cys34-Cys49, Cys39-Cys47), 2ILN (Cys9-Cys24, Cys12-Cys57, Cys14-Cys22, Cys31-Cys38, Cys35-Cys50, Cys40-Cys48), 1K9B (Cys3-Cys57, Cys4-Cys19, Cys7-Cys53, Cys9-Cys17, Cys27-Cys34, Cys31-Cys46, Cys36-Cys44), 1D6R (Cys3-Cys57, Cys4-Cys19, Cys7-Cys53, Cys9-Cys17, Cys27-Cys34, Cys31-Cys46, Cys36-Cys44), 6LH6 (Cys13-Cys66, Cys14-Cys29, Cys17-Cys62, Cys19-Cys27, Cys36-Cys43, Cys40-Cys55, Cys45-Cys53), 3MYW (Cys12-Cys66, Cys13-Cys28, Cys16-Cys62, Cys18-Cys26, Cys36-Cys43, Cys40-Cys55, Cys45-Cys53), 1H34 (Cys18-Cys72, Cys19-Cys34, Cys22-Cys68, Cys24-Cys32, Cys42-Cys49, Cys46-Cys61, Cys51-Cys59), 2G81 (Cys18-Cys72, Cys19-Cys34, Cys22-Cys68, Cys24-Cys32, Cys42-Cys49, Cys46-Cys61, Cys51-Cys59).
Figure S2. Example of amino acid sequences determination for native HOSPIs. The complete or a partial amino acid sequences of HOSPIs were assembled using the peptide sequences determined by Edman degradation, which were obtained by digestion of pyridylethylated HOSPIs peptides by each of the following enzymes (trypsin, Asp-N and Arg-C endoproteinases) or chemical breakdown (cyanogen bromide), respectively. Full length sequences were obtained for LOSPI-A4, -A6, -A7, -B1, -B4 to -B7, and -C1. Full length sequences were obtained for HOSPI-A4, -A6, -A7, -B1, -B4 to -B7, and -C1. The pyridylethylated protein after proteolytic digestion were separated from the reaction mixture by RP-HPLC on a MC-Pack ODS C₁₈ column (5 μm, 300 Å, 4.6 x 150 mm) using a linear gradient from 0% acetonitrile in 0.1% TFA to acetonitrile in 0.1% TFA. Each N-terminal sequence from obtained fragments is indicated as follows; P, the intact pyridylethylated protein, R; Arg-C, T; trypsin, E; Glu-C (V8 proteinase), D; Asp-N, M; cyanogen bromide. Cysteine residues in fragments are detected as a pyridylethylated cysteine.
**Figure S3.** Amino acid sequences of the linker region of HOSPIs. Asterisks indicate identical amino acids.
Figure S4. Expression of recombinant linker fusion and mature HOSPIs by E. coli. SDS-PAGE gels demonstrate HOSPI with linker (A) (pET32a-LOSPI-X1-lk) and without linker (B) (pET32a-LOSPI-X1-mat). Both proteins were expressed in Rosetta-gamiB (DE3) pLysS, as a fusion protein with thioredoxin-, His-tag, and S-tag derived of commercially available pET32a vector. The abbreviations used in the images are as follows; M, Marker; IPTG, isopropyl β-D-1-thiogalactopyranoside; IB, inclusion body; FT, flow-through, Wash, wash Ni²⁺-sepharose column factions, Elution, Ni²⁺-sepharose column elusion factions with high-concentration imidazole. Lanes of IPTG (-) and (+) correspond to the lysates of E.coli before and after induction of expression. Samples of other lanes correspond to each step of Ni²⁺ affinity purification after E.coli homogenization of the IPTG induces cultures. Finally, target proteins are obtained in elution fractions 1, 2, and 3. The bands of interest were detected approximately 32 and 30 kDa by SDS-PAGE analysis. (C) Final purification of pET32a-HOSPI-X1-lk by reversed-phase HPLC. The recombinant pET32a-HOSPI-X1-lk was eluted with a linear gradient of acetonitrile in 0.1% TFA (v/v) and monitored at 230 nm using a C₈ column (YMC pack C₈, 250×10 mm). Samples were lyophilized afterwards. SDS-PAGE image of Fr. 1 to 3 after HPLC analysis is shown as small inset in the upper left.
References of the supplementary file.

1. Park, E. Y., Kim, J. A., Kim, H. W., Kim, Y. S. and Song, H. K. (2004) Crystal structure of the Bowman-Birk inhibitor from barley seeds in ternary complex with porcine trypsin. J Mol Biol. 343, 173-186

2. Prakash, B., Selvaraj, S., Murthy, M. R., Sreerama, Y. N., Rao, D. R. and Gowda, L. R. (1996) Analysis of the amino acid sequences of plant Bowman-Birk inhibitors. J Mol Evol. 42, 560-569
**HOSPI-A2**  TWPPVEGRPSCKRC--GACTRx
**HOSPI-A3**  WPPVEGRPSCKRC--GACTRxx
**HOSPI-A4**  WPPVEGRPSCKRC--GACTRMWPÆANRCQQDDIVPO--CHEGQSKCEKDTRSGKLYQQOSFEYYNCAA
**HOSPI-A5**  WPPVEEGxFxxx
**HOSPI-A6**  WPPVEGRPCGRC--TACTLMLPDEANTCQQGVGDP--CHQQGSLCEVDPSPG---YQKXQESFYNGGKCP
**HOSPI-A7**  G^PyTWPPVEGRPCGRC--TACTLMLPDEANTCQQGVGDIVPK--CHQQGSCQKADTRSGPPLYQKLSFEYYNCAA
**HOSPI-B1**  GQQPLNCDC--FACARSL--CICQGDLVPO--CHEGQQCEFVT---GKLYQQOSFEYYNCA
**HOSPI-B3**  WPPVEEGPLCGRC--HACARMWPÆANRCQQGDxx
**HOSPI-B4**  WPPVEEGPLCGRC--HACARMWPÆANRCQQGDxx
**HOSPI-B5**  GRRPCCGC--GICORSLDP--HCQICEDLVPQ--CHEQCOACEEVDT---GTPMOYQRSFYHGCPTCL
**HOSPI-B6**  GRRPGGCNC--GICORSLDP--HCQICEDLVPQ--CHEQCOACEEVDT---GTPMOYQRSFYHGCPTCL
**HOSPI-B7**  GEGACGNCNDCFGGRAGFS---RCROCLDLVVK--CHPSCSNCMTETYPWYQGQPLMDDDCATP
**HOSPI-C1**  E-RPLCNEC--FICORSGDPO--RCLQEDPQVPO--CHEQCOKEKDRTSGETMYQRSFYHDCANE

**OnionBBI**  GDEEEGCGGGSC--GSDQRADLA--RQECDRVTSCGPGKCRCEAADLNNPRVYKOMSHCSCOQCRSIL
**WheatBBI**  EEAMPSAWPCGDC--GTCTRMIPP---RCROMDVSPGSGPACNKQVOTTILGGRDVFVW--GMLRIENFCKRQGTPAR
**SoybeanBBI**  DDESSKPGCQDC--ACTKSNPP---RQRSGMRLNSCHSACKSCALSYPAQ---FQVIDTFQCYEOPKPSEDDKEN

---

**First reactive site loop**

---

**Second reactive site loop**

(for soybeanBBI only)
(A)

HOSPI-1

HOSPI-2-6

HOSPI-x3

HOSPI-1-3

HOSPI-1-8

HOSPI-2-4

(B)
